The Astronomical Journal, 155:89 (24pp), 2018 February https://doi.org/10.3847/1538-3881/aaa54c

The California-*Kepler* Survey. IV. Metal-rich Stars Host a Greater Diversity of Planets

Erik A. Petigura[1,7], Geoffrey W. Marcy[2], Joshua N. Winn[3], Lauren M. Weiss[4,8], Benjamin J. Fulton[1], Andrew W. Howard[1], Evan Sinukoff[5], Howard Isaacson[2], Timothy D. Morton[4], and John Asher Johnson[6]
[1] California Institute of Technology, Pasadena, CA 91125, USA; petigura@caltech.edu
[2] Department of Astronomy, University of California, Berkeley, CA 94720, USA
[3] Princeton University, Princeton, NJ 08544, USA
[4] University of Montreal, Montreal, QC, H3T 1J4, Canada
[5] Institute for Astronomy, University of Hawai'i at Mānoa, Honolulu, HI 96822, USA
[6] Harvard-Smithsonian Center for Astrophysics, Cambridge, MA 02138, USA
Received 2017 October 24; revised 2017 December 23; accepted 2017 December 27; published 2018 January 29

Abstract

Probing the connection between a star's metallicity and the presence and properties of any associated planets offers an observational link between conditions during the epoch of planet formation and mature planetary systems. We explore this connection by analyzing the metallicities of *Kepler* target stars and the subset of stars found to host transiting planets. After correcting for survey incompleteness, we measure planet occurrence: the number of planets per 100 stars with a given metallicity M. Planet occurrence correlates with metallicity for some, but not all, planet sizes and orbital periods. For warm super-Earths having $P = 10$–100 days and $R_P = 1.0$–$1.7\ R_\oplus$, planet occurrence is nearly constant over metallicities spanning -0.4 to $+0.4$ dex. We find 20 warm super-Earths per 100 stars, regardless of metallicity. In contrast, the occurrence of warm sub-Neptunes ($R_P = 1.7$–$4.0\ R_\oplus$) doubles over that same metallicity interval, from 20 to 40 planets per 100 stars. We model the distribution of planets as $df \propto 10^{\beta M} dM$, where β characterizes the strength of any metallicity correlation. This correlation steepens with decreasing orbital period and increasing planet size. For warm super-Earths $\beta = -0.3^{+0.2}_{-0.2}$, while for hot Jupiters $\beta = +3.4^{+0.9}_{-0.8}$. High metallicities in protoplanetary disks may increase the mass of the largest rocky cores or the speed at which they are assembled, enhancing the production of planets larger than $1.7\ R_\oplus$. The association between high metallicity and short-period planets may reflect disk density profiles that facilitate the inward migration of solids or higher rates of planet–planet scattering.

Key words: methods: statistical – planets and satellites: formation – planets and satellites: general – stars: abundances – stars: fundamental parameters – techniques: spectroscopic

Supporting material: machine-readable tables

1. Introduction

Exploring the connection between planets and their host stars has been a long-standing focus of exoplanet astronomy. Host star metallicity is thought to reflect the metallicity of the protostellar nebula and the protoplanetary disk from which planets form. Metal-rich protoplanetary disks are thought to have enhanced surface densities of solids. Viewed in the context of core-accretion theory (Lissauer 1995; Pollack et al. 1996), one might expect metal-rich disks to form terrestrial planets and the cores of gas giant planets with greater efficiency than metal-poor disks. If true, metal-rich stars should host greater numbers of gas giant and terrestrial planets. This prediction can be tested by studying the correlation (or lack thereof) between [Fe/H] and planet occurrence.

The extent to which stellar metallicity correlates with the presence or absence of planets has been the subject of many previous studies. Gonzalez (1997) observed that the first four extrasolar planets discovered orbited metal-rich stars and concluded that metal-rich stars have metal-rich protoplanetary disks which form planets more efficiently. As the sample of Doppler-detected planets grew into the hundreds, various studies noted a preference for Jovian-mass planets to orbit stars with super-solar metallicities (e.g., Santos et al. 2004; Fischer & Valenti 2005). However, as Doppler surveys pushed into lower regimes of planet mass, various authors noted that the correlation between planet occurrence and host star metallicity appeared to weaken (e.g., Sousa et al. 2008; Ghezzi et al. 2010).

The prime *Kepler* mission (2009–2013) revealed more than 4000 planet candidates with sizes as small as Mercury (Barclay et al. 2013). In contrast to previous studies that explored the connection between host star metallicity and planet mass, *Kepler* studies have focused mainly on the connection between metallicity and planet size. While a handful of *Kepler* planets have well-measured masses through radial velocities (RVs; e.g., Marcy et al. 2014) or transit-timing variations (TTVs; e.g., Hadden & Lithwick 2017). The vast majority of *Kepler* planets have unknown masses. Current RV instruments require bright targets ($V < 13$ mag), and TTV measurements require tightly packed, multi-planet systems. Both TTV and RV mass measurements are only possible on a small fraction of the total *Kepler* planet sample.

While the *Kepler* sample provided a large sample of planets for planet-metallicity studies, extensive follow-up spectroscopy was first required to measure host star metallicities. Buchhave et al. (2012) measured the metallicities of 152 stars harboring 226 planets and observed that while planets larger than $4\ R_\oplus$ orbit metal-rich stars, smaller planets orbit stars with wide-ranging metallicities. Later, using an augmented sample of 405 stars hosting 600 planets, Buchhave et al. (2014) argued for three distinct stellar metallicity distributions, with breakpoints

[7] Hubble Fellow.
[8] Trottier Fellow.

at 1.7 $R_\oplus$ and 3.9 $R_\oplus$. In contrast, Schlaufman (2015) found no evidence for different metallicity distributions above and below 1.7 $R_\oplus$ and raised several concerns regarding the statistical validity of the Buchhave et al. (2014) analysis.

RV mass measurements of transiting planets have revealed that planets smaller than 1.7 $R_\oplus$ have bulk densities consistent with rocky compositions (e.g., Weiss & Marcy 2014; Rogers 2015), at least for the short-period planets ($P \lesssim 20$ days) that are amenable to these RV mass measurements. Consequently, the degree to which metallicity correlates with the occurrence of planets smaller than 1.7 $R_\oplus$ is of particular interest. On this point, there is disagreement in the literature. Wang & Fischer (2015), using low precision metallicities from the *Kepler* Input Catalog (KIC; Brown et al. 2011), reported that the rate of planets smaller than 1.7 $R_\oplus$ is $1.72^{+0.18}_{-0.17}$ times higher for stars with [Fe/H] > 0 dex compared to stars [Fe/H] < 0 dex. In contrast, Buchhave & Latham (2015) found no evidence of a metallicity enhancement among stars hosting planets smaller than 1.7 $R_\oplus$.

A long-standing limitation of *Kepler* metallicity studies was that reliable metallicities did not exist for a representative number of *Kepler* field stars. In constructing the KIC, Brown et al. (2011) placed a prior on the *Kepler* field metallicities based on the metallicities of nearby stars, as measured by Nordström et al. (2004), which have a mean of −0.14 dex and dispersion of 0.19 dex. However, it was not clear whether *Kepler* field stars, which are typically ∼1 kpc from Earth, would follow the metallicity distribution of nearby stars.

Such metallicity offsets have been invoked to explain differences in planet occurrence rates between the *Kepler* field and the solar neighborhood. Howard et al. (2012) measured a hot Jupiter occurrence rate of 0.4 ± 0.1%, roughly 40% of that in the solar neighborhood (1.2 ± 0.4%, Wright et al. 2012), and speculated that "a paucity of metal-rich stars in the *Kepler* sample is one possible explanation."

Today we know that *Kepler* field stars have a *higher* mean metallicity than solar neighborhood stars. The Large Sky Area Multi-object Fibre Spectroscopic Telescope (LAMOST; Cui et al. 2012; Luo et al. 2012; Zhao et al. 2012) is instrumented with a multiplexed, low-resolution spectrometer (4000 fibers, $R = 1800$) and can therefore efficiently gather spectra for large samples of *Kepler* target stars with and without transiting planets. Dong et al. (2014), through analyzing the LAMOST Data Releases 1 and 2, found that the mean metallicity of 12,000 *Kepler* field stars was −0.04 dex, much closer to solar than to the value of −0.14 dex assumed in the construction of the KIC. Guo et al. (2017) also found a near-solar mean metallicity of −0.04 dex in an analysis of 610 *Kepler* field stars observed by the Hectochelle $R = 34{,}000$ spectrometer at the MMT. Thus, a metallicity offset cannot explain differences in the hot Jupiter rates.

The LAMOST data sets permitted several breakthroughs in *Kepler* planet-metallicity studies, by measuring the true metallicity distribution of bright ($Kp < 14$) *Kepler* field stars. Mulders et al. (2016) analyzed the LAMOST metallicities and a sample of 665 planet candidates, and found that the occurrence rate of hot small planets ($P < 10$ days, $R_P < 4\,R_\oplus$) is three times higher among super-solar metallicity hosts compared to sub-solar hosts. Dong et al. (2018) also analyzed LAMOST metallicities and a sample of 295 planets and reported a similar trend, also noting that hot Neptune-size planets are typically single.

Here, we examine the connection between planets and stellar metallicity using spectroscopy from the California-*Kepler* Survey (CKS). Given that the CKS produced a homogeneous set of high-precision metallicities of 1305 stars, we can explore this connection in unprecedented detail. A key advantage of the CKS sample is the highly precise planet radii (10% precision) and stellar metallicities (0.04 dex precision) compared to previous studies. The CKS sample has high purity (i.e., low false positive rate) due to extensive vetting of false positives. This data set has already revealed new features in the planet radius distribution, most notably a gap in the size distribution of small ($R_P = 1$–$4\,R_\oplus$ planets) reported by Fulton et al. (2017).

We find that planets larger than Neptune are preferentially found around metal-rich stars, while planets smaller than Neptune are found around stars of wide-ranging metallicity. For super-Earth-size planets ($R_P = 1.0$–$1.7\,R_\oplus$), we observe a positive metallicity correlation for $P = 1$–10 days, but no correlation for $P = 10$–100 days. In contrast, rates of sub-Neptune-size planets ($R_P = 1.7$–$4.0\,R_\oplus$) correlate with metallicity over $P = 1$–100 days. Planets larger than Neptune are an order of magnitude less common than planets smaller than Neptune, and probing the possible metallicity correlations is more challenging. However, we observe strong metallicity correlations for both Jovian-size ($R_P = 8.0$–$24.0\,R_\oplus$) and sub-Saturn-size ($R_P = 4.0$–$8.0\,R_\oplus$) planets.

We describe our planet and stellar samples in Section 2. Section 3 explores the distribution of planets in the P–R_P plane as a function of metallicity, and highlights areas where metallicity plays a strong effect. In Sections 4–6, we compare host star metallicities to the *Kepler* field star distribution and compute planet occurrence as a function of metallicity. We summarize our findings in Section 7, and offer some interpretations of the observed trends.

2. Sample

Studies of planet occurrence require both a sample of planets $\mathcal{P}$ and a parent stellar sample $\mathcal{S}$ from which the planets are drawn. We construct $\mathcal{P}$ from the CKS sample (Section 2.1) and apply a series of filters aimed at creating a well-defined sample of high purity with well-measured radii (Section 2.2). We then construct $\mathcal{S}$ from the *Kepler* field stars after applying the same set of filters used to construct $\mathcal{P}$ (Section 2.3). Because the metallicities are not known for every star in $\mathcal{S}$, we employ the LAMOST parameters as a proxy (Section 2.4).

2.1. Initial Planet Sample

CKS is a large-scale spectroscopic survey of 1305 *Kepler* Objects of Interest (KOIs). The sample selection, spectroscopic observations, and spectroscopic analysis are described in detail in Petigura et al. (2017b, hereafter Paper I). In brief, the sample was initially constructed by selecting all KOIs brighter than $Kp = 14.2$ mag.[9] A KOI is a *Kepler* target star which showed periodic photometric dimmings indicative of planet transits. However, not all KOIs have received the necessary follow-up attention needed to confirm the planets.

Over the course of the project, we included additional targets to cover different planet populations, including multi-candidate hosts, Ultra Short Period candidates, and Habitable Zone

[9] *Kepler* magnitude; $Kp \approx V - 0.4$ mag for G2 stars.

candidates. We obtained spectra at the 10 m Keck Telescope using the High Resolution Echelle Spectrometer at a resolution of $R = 60{,}000$ (Vogt et al. 1994). A key feature of the CKS data set is that stars were observed to a consistent signal-to-noise ratio (S/N) of 45 pixel^{-1} at the peak of the blaze function near 5500 Å.

In Paper I, we extracted the following properties for each star: effective temperature $T_{\rm eff}$, surface gravity $\log g$, metallicity [Fe/H], and projected rotational velocity $v \sin i$ using the SpecMatch (Petigura 2015) and SME@XSEDE codes (Paper I).

In Johnson et al. (2017, hereafter Paper II), we converted the spectroscopic properties of $T_{\rm eff}$, $\log g$, and [Fe/H] into stellar mass $M_\star$, radius $R_\star$, and age. This conversion is facilitated by the publicly available `isochrones` Python package (Morton 2015). Stellar mass, radius, and age are measured to 4%, 11%, and 30%, respectively. We then recompute planetary radii and equilibrium temperatures using our updated CKS parameters and the results from transit fitting performed by Mullally et al. (2015). Paper II provides updated stellar and planetary properties for 1305 KOIs and 2025 planet candidates, which are the starting point for our planet sample $\mathcal{P}$.

2.2. Filtered Planet Sample

We applied a series of filters, described below in bullets, to the CKS sample to arrive at our final planet sample $\mathcal{P}$. The filters restrict the range of stellar properties included in our analysis. Since we aim to explore the connection between metallicity and planet size, we also require the planets have well-determined radii.

Where possible, we filter based on the DR25 stellar properties table of Mathur et al. (2017). These cuts may be applied homogeneously to the field star sample $\mathcal{S}$. Some filters involve follow-up observations not available for every star in $\mathcal{S}$. In Section 2.3, we quantify the number of field stars that would have been excluded had comparable follow-up been performed. The applied filters are as follows:

1. *Stellar brightness.* We restrict our sample to the magnitude-limited sub-sample of the CKS sample (i.e., $Kp < 14.2$ mag).
2. *Stellar effective temperature*. The spectroscopic tools used in Paper I produce reliable results for $T_{\rm eff} =$ 4700–6500 K. We restrict our analysis to stars having photometric $T_{\rm eff} =$ 4700–6500 K (DR25 stellar properties table).
3. *Stellar surface gravity*. We restrict our analysis to stars having photometric $\log g =$ 3.9–5.0 dex (DR25 stellar properties table).
4. *Stellar dilution*. Dilution from nearby stars can also alter the apparent planetary radii. Furlan et al. (2017) compiled high resolution imaging observations performed by several groups.[10] When a nearby star is detected, Furlan et al. (2017) computed a radius correction factor (RCF), which accounts for dilution assuming the planet transits the brightest star. We elect to not apply this correction factor, but conservatively exclude KOIs where the RCF is larger than 5%.
5. *Planet orbital period*. We remove planet candidates with orbital periods longer than 350 days. This excludes planets that only transit once or twice during *Kepler* observations, which have a higher false positive rate (Mullally et al. 2015).
6. *Planet false positive designation*. We exclude candidates that are identified as false positives according to Paper I.
7. *Planets with grazing transits*. Finally, we exclude stars having grazing transits ($b > 0.9$), which have suspect radii due to covariances with the planet size and stellar limb-darkening during the light curve fitting.

[10] Alphabetical by author: Adams et al. (2012, 2013), Baranec et al. (2016), Cartier et al. (2015), Dressing et al. (2014), Everett et al. (2015), Gilliland et al. (2015), Horch et al. (2012, 2014), Howell et al. (2011), Law et al. (2014), Lillo-Box et al. (2012, 2014), Wang et al. (2015a, 2015b).

Table 1
Filters Applied to Planet Sample

Filter	$n_{\rm pl,pass}$	$n_{\rm pl,pass,run}$	$f_{\rm pl,pass,run}$	$n_{\star,\rm pass,run}$
Full sample	2025	2025	1.000	1279
$Kp < 14.2$ mag	1359	1359	0.671	954
$T_{\rm eff} =$ 4700–6500 K	1977	1328	0.977	929
$\log g =$ 3.9–5.0 dex	1899	1212	0.913	829
$P < 350$ days	1965	1191	0.983	817
Not a false positive	1861	1105	0.928	758
Radius correction factor <5%	1891	1017	0.920	695
Not a grazing transit ($b < 0.9$)	1870	970	0.954	662

Note. Summary of the filters applied to the CKS catalog to create planet sample $\mathcal{P}$. The column labeled $n_{\rm pl,pass}$ is the total number of KOIs that pass a specific filter and $n_{\rm pl,pass,run}$ is a running tally of KOIs that pass all filters. For filter i, $f_{\rm pl,pass,run}(i) = n_{\rm pl,pass,run}(i)/n_{\rm pl,pass,run}(i-1)$. For example, $f_{\rm pl,pass,run}(3) = n_{\rm pl,pass,run}(3)/n_{\rm pl,pass,run}(2) = 1328/1359 = 0.977$ The column labeled $n_{\star,\rm pass,run}$ gives the numbers of unique stars hosting the $n_{\rm pl,pass,run}$ KOIs.

Our successive filters, along with a running tally of the number planets which pass them, are listed in Table 1. The filters on dilution and grazing transits each eliminate a small percentage (8% and 5%) of planets with imprecise radii, even after CKS spectroscopy. Similar filters clarified the bimodal planet radius distribution presented by Fulton et al. (2017), Paper III, in the CKS series.

In total, $\mathcal{P}$ contains 970 planets orbiting 662 stars that pass all filters. Figure 1 shows planets on the [Fe/H]–R_P plane as successive cuts are applied. Even though planets across this plane are filtered out, planets larger than Neptune have a higher rate of removal. Properties of the stars hosting these planets are shown in Figure 2.

2.3. Field Star Sample

In order to compare the properties of stars with and without transiting planets, we need to consider the parent population of *Kepler* field stars $\mathcal{S}$, from which $\mathcal{P}$ is drawn. We begin with the Mathur et al. (2017) catalog that lists all 199,991 stars observed at some point during the *Kepler* mission. Where possible, we apply the same set of filters to construct $\mathcal{S}$ that were used to construct $\mathcal{P}$. Table 2 summarizes the number of stars that pass cuts on Kp, photometric $T_{\rm eff}$, and photometric $\log g$. A total of 36,959 stars pass all cuts. The distribution of field star properties is shown in Figure 2.

The filters used to construct $\mathcal{P}$ were not simply cuts on the stellar properties, but also on quality of the CKS stellar radii and the properties of the planet candidates. We cannot apply these same filters to the field stars on a star-by-star basis

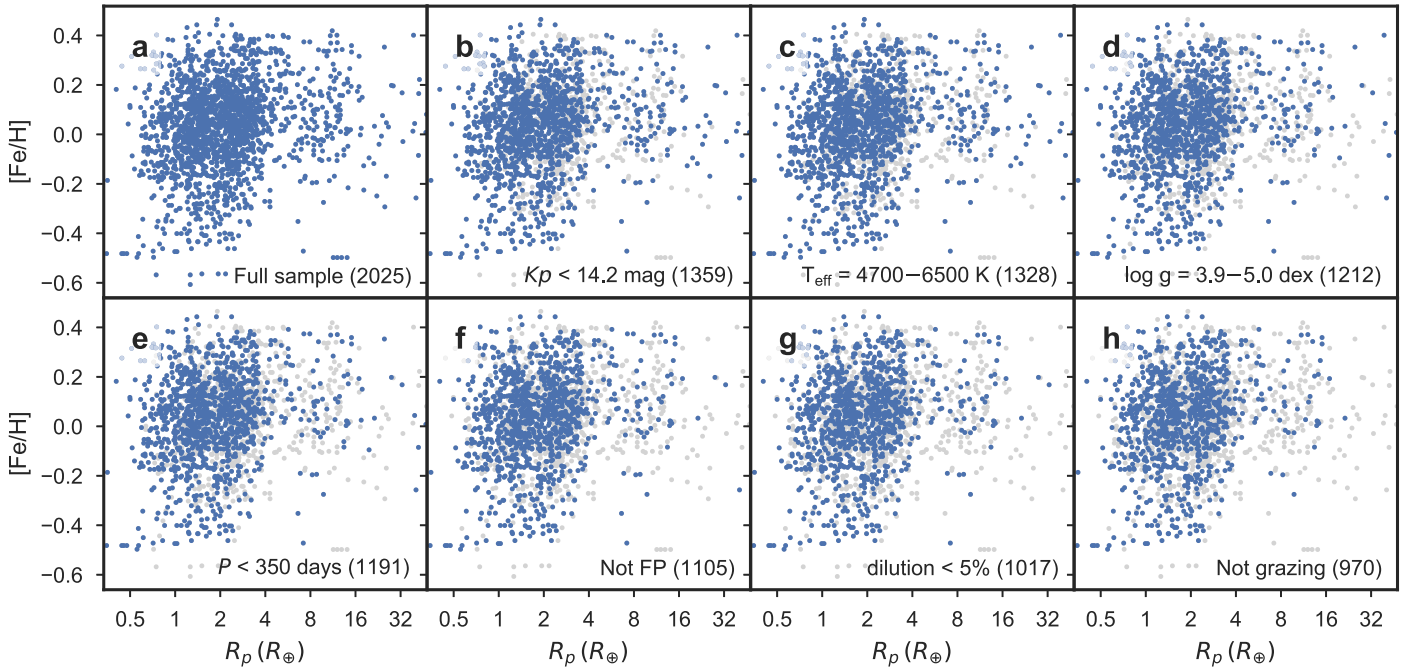

Figure 1. Host star metallicity vs. planet size after the application of successive filters. Panel (a): all KOIs in the CKS sample. Panels (b)–(j): blue points show the KOIs that pass successive filters in stellar brightness, effective temperature, surface gravity, planetary orbital period, false positive disposition, radius precision, dilution due to nearby stars, and impact parameter. The gray points show KOIs that do not pass one or more of the filters.

because these stars do not have CKS stellar radii or detected planets. However, we must assess whether the remaining filters from Section 2.2 would have excluded a significant fraction of field stars, assuming all stars received similar follow-up attention. Here, we quantify the number of field stars that would have been excluded from $\mathcal{S}$ had comparable follow-up been performed.

1. *Reliable spectroscopic parameters.* Our initial sample of planets orbit stars for which the spectroscopic analysis described in Section 2 produced reliable results. For a star to be included, $v \sin i$ must be less than 20 km s^{-1}. This selection effect excludes some stars that are typically near 6500 K, where $v \sin i$ values begin to exceed 20 km s^{-1}. After filtering on *Kp*, photometric $T_{\rm eff}$, and photometric $\log g$, 3.1% of stars are excluded because they do not have reliable spectroscopic parameters.
2. *Stellar dilution*. 7.8% of planet candidates were excluded because the *Kepler* apertures contained enough flux from neighboring stars, such that the planet radii required a correction factor larger than 5%. The fraction of stars excluded depends on the crowding of *Kepler* field stars. We make the assumption that this distribution is independent of KOI status. Had all field stars received a similar level of high-contrast imaging follow-up, 7.8% would have companions sufficiently bright and close to warrant a RCF of >5%.
3. *Planet orbital period*. For a KOI to be included in our sample, we required that the orbital period be less than 350 days. Such a cut is strictly a cut on the planet properties and does not preclude any stars from being included in $\mathcal{S}$.
4. *Planet false positive disposition*. A star must first be identified as a KOI in order to then be designated as a false positive. Therefore, the false positive filter does not exclude significant numbers of field stars from $\mathcal{S}$.
5. *Planets with grazing transits*. We require that the planets have non-grazing impact parameters. As with the radius filter, it does not affect inclusion in $\mathcal{S}$. However, this cut on impact parameter must be factored into the geometric transit probability, which affects the occurrence calculation described in Section 4.

When we account for stars that would have been excluded from our sample due to unreliable spectroscopic parameters or the presence of nearby stars, we find that the planet sample $\mathcal{P}$ was drawn from a parent stellar sample $\mathcal{S}$ containing 36,959 × 0.922 × 0.969 = 33,020 stars.

2.4. Metallicity Distribution of Kepler Field Stars

Here, we characterize the metallicity distribution of the parent sample $\mathcal{S}$. While the KIC (Brown et al. 2011) and its updates (e.g., Mathur et al. 2017) tabulate metallicities for every *Kepler* target star, they are inadequate for characterizing the metallicity distribution of the *Kepler* field given their low precision. Instead, we use the LAMOST/LASP stellar parameters (Luo et al. 2015) from Data Release 3 (DR3).[11] Following De Cat et al. (2015), we crossmatch the LAMOST-DR3 with the KIC by identifying LAMOST spectra taken within 1.2 arcsec of a KIC target star. This crossmatching results in 29,997 stars in common having $Kp < 14.2$ mag. We note a systematic offset between the LAMOST and CKS metallicity scales of ≈0.04 dex when comparing 476 stars in common. We apply a correction, described in Appendix A, but estimate that residual systematic offsets of ≈0.01 dex remain.

We apply the same set of filters to the LAMOST stars that we applied to $\mathcal{P}$. These include cuts in *Kp*, $T_{\rm eff}$, and $\log g$. After

[11] http://dr3.lamost.org

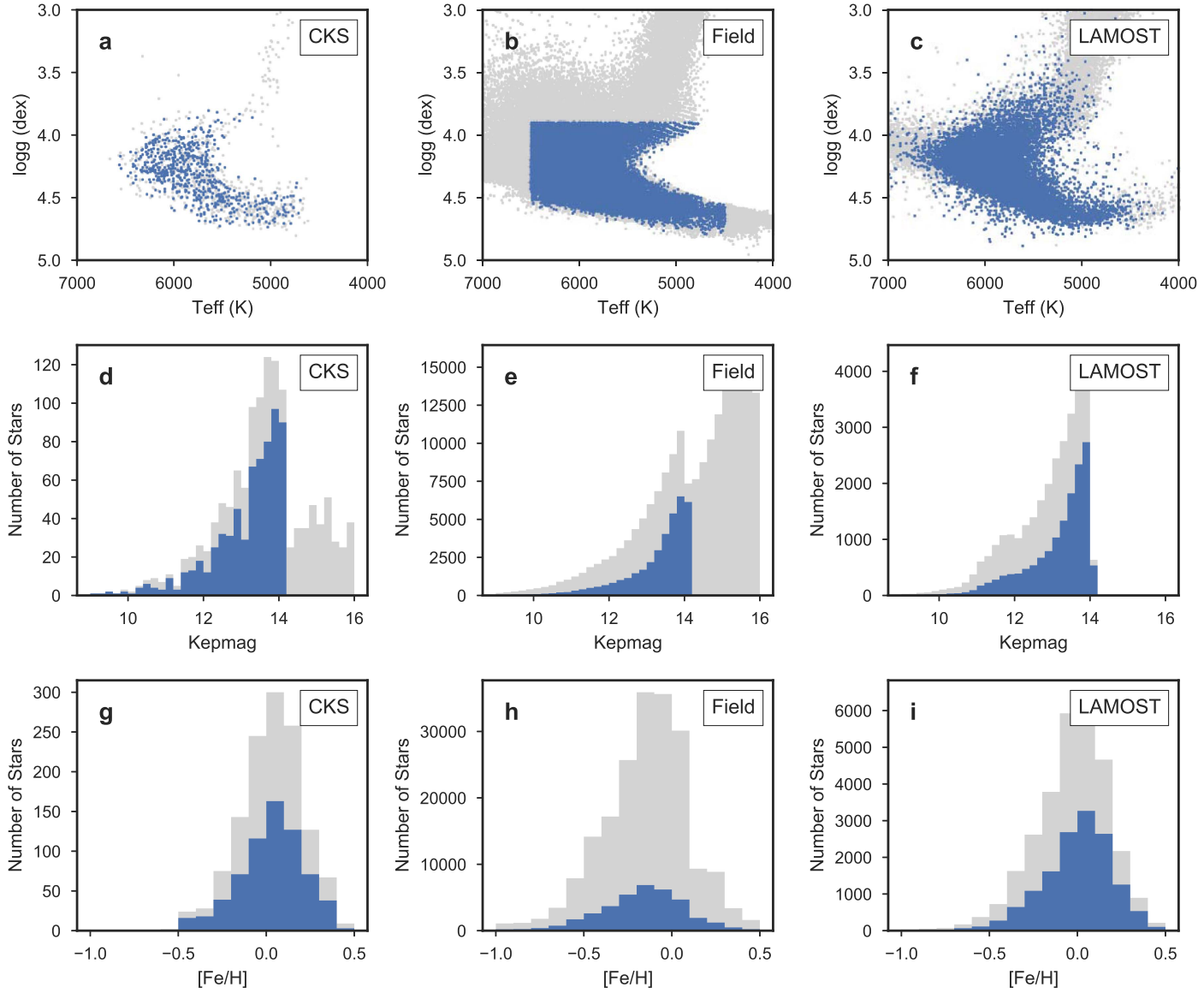

Figure 2. The distribution of stellar properties for the three samples of stars considered in this work. Panel (a): blue points show $T_{\rm eff}$ and $\log g$ of CKS planet hosts that passed all filters $\mathcal{P}$; gray points represent all CKS planet hosts. Panels (b) and (c) show the same quantities as panel (a), but for the *Kepler* target stars and the LAMOST sample, respectively. Panels (d)–(f): distributions of host star *Kp*. Panels (g)–(i): distributions of host star metallicity from different catalogs. The sub-solar mean metallicity of the *Kepler* target stars (h) is a reflection of the low metallicity prior used in the photometric modeling.

Table 2
Filters Applied to Stellar Sample

Cut	$n_{\star,\rm pass}$	$n_{\star,\rm pass,run}$	$f_{\star,\rm pass,run}$
Full sample	199991	199991	1.000
$Kp < 14.2$ mag	81758	81758	0.409
$T_{\rm eff} = 4700$–6500 K	168885	62751	0.768
$\log g = 3.9$–5.0 dex	162854	36959	0.589

Note. Summary of the filters applied to $\mathcal{S}$. See Table 1 for column descriptions.

Table 3
Summary of Cuts to LAMOST Sample

Cut	$n_{\star,\rm pass}$	$n_{\star,\rm pass,run}$	$f_{\star,\rm pass,run}$
Full sample	29997	29997	1.000
$Kp < 14.2$ mag	29997	29997	1.000
$T_{\rm eff} = 4700$–6500 K	23551	23551	0.785
$\log g = 3.9$–5.0 dex	18625	14382	0.611

Note. Summary of the cuts applied to the LAMOST DR-2 sample. See Table 1 for column descriptions.

applying these cuts, we are left with 14,382 stars with LAMOST parameters. The filters are summarized in Table 3 and the distribution of LAMOST stellar properties is shown in Figure 2.

Table 4 summarizes the metallicity distribution of $\mathcal{P}$ and $\mathcal{S}$ measured through different methods. The mean metallicity of $\mathcal{S}$, as measured by LAMOST, is -0.01 dex, similar to the mean metallicity of the filtered planet sample $\mathcal{P}$, $+0.03$ dex, as measured by CKS. We note that the mean metallicity of $\mathcal{S}$, according to the DR25 stellar properties table (Mathur et al. 2017), is -0.19 dex. This low value is due to the low metallicity prior used in the photometric modeling.

Table 4
Comparison of Planet Host and Field Star Metallicities

	$\mathcal{P}$ (spec) (dex)	$\mathcal{S}$ (phot) (dex)	$\mathcal{S}$ (spec) (dex)
Mean	0.031	−0.187	−0.005
rms	0.185	0.259	0.207
SEM	0.006	0.001	0.002
25%	−0.069	−0.320	−0.116
50%	0.052	−0.160	0.020
75%	0.148	−0.020	0.131

Note. Summary of metallicity distributions for different stellar samples. $\mathcal{P}$ (spec) refers to the CKS metallicities of our filtered list of planet hosts (Section 2.2). $\mathcal{S}$ (phot) refers to the photometric metallicities of our parent stellar sample (Section 2.3). $\mathcal{S}$ (spec) refers to the metallicities of $\mathcal{S}$ measured using LAMOST data (Section 2.4). $\mathcal{S}$ (phot) and $\mathcal{S}$ (spec) have significantly different mean metallicities, due the metallicity prior imposed by Mathur et al. (2017).

3. Metallicities of Planet Hosts

In this section, we examine the metallicities of planet host stars with respect to planet size and orbital period. We first briefly note the trends seen within the planet sample $\mathcal{P}$, without reference to field star metallicities (Section 3.1). However, features in P–R_P–[Fe/H] distribution are much more apparent when compared to the metallicities of the parent stellar sample $\mathcal{S}$. We perform this analysis in Section 3.2. By searching for evidence of elevated metallicity in the host stars of certain types of planets, we can identify planet classes that show some positive correlation with metallicity. The advantage of this approach is that it does not require any modeling of the *Kepler* survey completeness. A more difficult but ultimately more useful approach is to compute planet occurrence as a function of metallicity, which requires modeling survey completeness. This approach is taken in Section 6.

3.1. Properties of Planet Hosts

In Figure 3, we show the planet sizes and host star metallicities for the 2025 planet candidates in the CKS sample. Planets smaller than Neptune are found around stars of wide-ranging metallicities, while there is a deficit of planets larger than Neptune around stars with sub-solar metallicity.

To investigate variation in average host star metallicities with planet sizes, we divided the metallicity measurements according to planet size. Bins span a factor of $\sqrt{2}$ in R_P for planets smaller than 4 $R_\oplus$. Larger planets are placed into bins spanning a factor of 2, owing to lower numbers. Figure 3 shows the mean metallicities for these planet radius bins along with the 25% and 75% quantiles. We observe a gradual upward trend in mean host star metallicity from smaller to larger planets. Buchhave et al. (2012, 2014) observed a similar trend in smaller samples of planet hosts.

Figure 3 also shows host star metallicity as a function of orbital period. Unlike the R_P–[Fe/H], there are no large regions of the P–[Fe/H] plane clearly devoid of planets. After computing the mean metallicity in bins of P spanning 0.25 dex, we observe a slight increase in mean metallicity of about 0.05 dex with decreasing orbital period over P = 1–10 days. Mulders et al. (2016) observed a similar trend in a smaller sample of planet hosts.

3.2. Comparison to Field Stars

Here, we examine the effect of host star metallicity on the 2D distribution of planet size and orbital period. We divided $\mathcal{P}$ into four bins of host star metallicity with boundaries at −0.116, +0.020, and +0.131 dex. The boundaries of the bins equally divide the stars in $\mathcal{S}$ (see Table 4). Figure 4 shows the distribution of planets in the P–R_P plane for the different metallicity bins. If the occurrence of planets were independent of host star metallicity, then the distribution of planets in each plot would be indistinguishable. The panels of Figure 4 show clear differences, indicating that metallicity is associated with certain types of planets, and that the strength of that enhancement depends on both P and R_P.

To facilitate our investigation into the effect of metallicity across the P–R_P plane, we define several regions. As a matter of convenience, we define a nomenclature for referring to these different regions. While the boundaries are matters of taste, we choose physically motivated boundaries, when possible. We consider four domains of planet size, defined below:

1. *Jupiters*. R_P = 8–24 $R_\oplus$. The lower limit is motivated by the fact that planets larger than 8 $R_\oplus$ tend to have masses ranging from 100 to 10,000 $M_\oplus$, while planets smaller than 8 $R_\oplus$ tend to have lower masses ranging from 6 to 60 $M_\oplus$ (see, e.g., Petigura et al. 2017b, Figure 9). At $M_P \approx$ 100 $M_\oplus$ electron degeneracy pressure begins contribute significantly to a planet's pressure support (Zapolsky & Salpeter 1969). Thus 8 $R_\oplus$ approximates the dividing line between planets with different pressure support and interior structures. The upper radius limit of 24 $R_\oplus$ is based on the known sizes of hot Jupiters. The most highly irradiated Jovians discovered to date can exceed 2 $R_{\rm jup}$ (e.g., WASP-79b; Smalley et al. 2012).[12,13]
2. *Sub-Saturns*. R_P = 4–8 $R_\oplus$. Sub-Saturns are typically typically less massive the larger Jupiters. They are roughly 10× more rare than the smaller sub-Neptunes, suggesting a different formation pathway. A radius of 4 $R_\oplus$ approximates a breakpoint in the planet size distribution where planet occurrence rises rapidly with decreasing size (see Fulton et al. 2017).
3. *Sub-Neptunes*. R_P = 1.7– 4.0 $R_\oplus$. The lower radius limit corresponds to a likely transition radius between rocky planets and planets that have envelopes that are an appreciable fraction of the total planet size. Among the observations supporting such a transition are measurements of planet bulk density from transits and radial velocities (Marcy et al. 2014; Weiss & Marcy 2014; Rogers 2015) and the observation of a gap in the radius distribution of planets by Fulton et al. (2017).
4. *Super-Earths*. R_P = 1.0–1.7 $R_\oplus$. Planets that are smaller than sub-Neptunes. Few planets in this size range have well-measured masses, but those that do are often consistent with rocky compositions.

We also consider three domains of orbital period:

1. *Hot Planets*. P = 1–10 days. The upper limit of 10 days corresponds to a breakpoint in distribution of planet occurrence with orbital period (Howard et al. 2012; Fressin et al. 2013). Below $P \approx$ 10 days, planet

[12] Exoplanet Archive (Akeson et al. 2013).
[13] https://exoplanetarchive.ipac.caltech.edu/

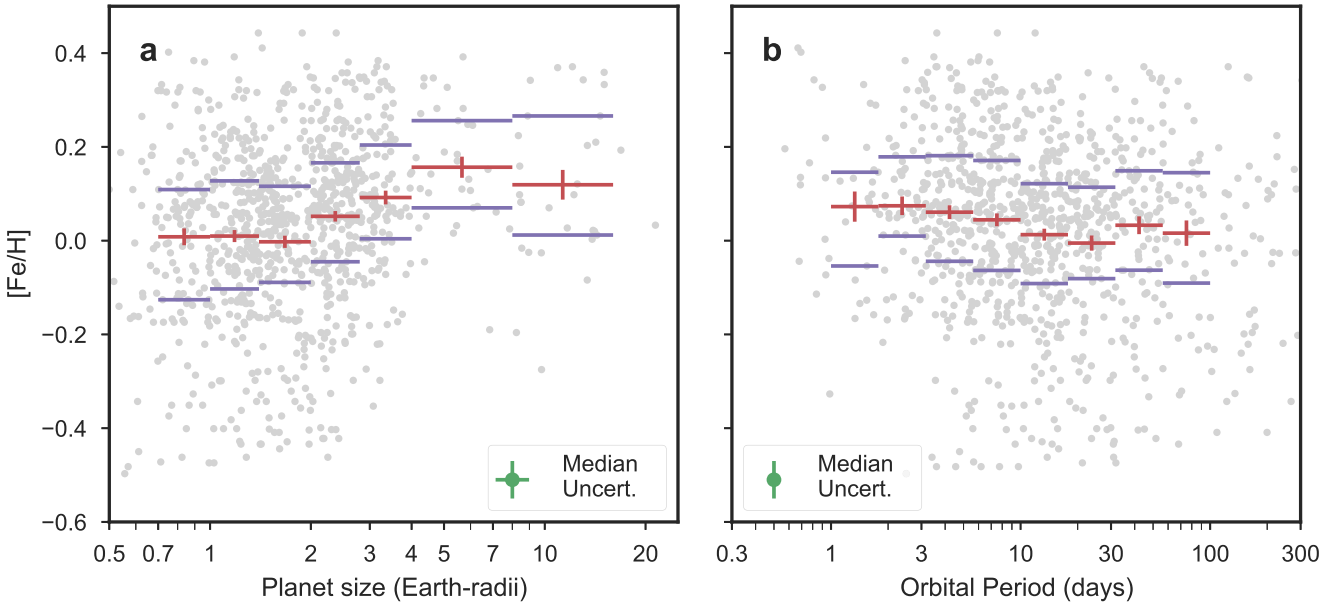

Figure 3. Panel (a) shows the sizes and host star metallicities for the 970 planets in the filtered CKS planet catalog $\mathcal{P}$. We observe a clear deficit of planets larger than Neptune with sub-Solar metallicities. The Fulton et al. (2017) radius gap may be seen from $R_P = 1.5$–$2.0\ R_\oplus$. We show the mean host star metallicity for various size ranges of planets with the red lines. The vertical bars show the standard error of the mean. The purple lines show the 25% and 75% quantiles. Mean metallicity is roughly constant from 0.7 $R_\oplus$ to 2.0 $R_\oplus$, rises from 2.0 $R_\oplus$ to 4.0 $R_\oplus$, and is roughly constant from 4 $R_\oplus$ to 16 $R_\oplus$. Panel (b): same as (a) except showing orbital period on the x-axis. We observe a small 0.05 dex increase in mean metallicity with decreasing orbital period over $P = 1$–10 days.

occurrence per $\log P$ is observed to decline, while planet occurrence is roughly constant for longer periods.

2. *Warm Planets.* $P = 10$–100 days. An intermediate range of orbital periods. While warm planets are intrinsically more common than hot planets, they represent about half of the total sample due to falling completeness and transit probability with increasing orbital period.
3. *Cool Planets.* $P = 100$–350 days. The longest period planets included in our survey. There are very few planets in our sample with such long periods due to falling completeness and decreasing transit probability.

Here, we note some qualitative features in Figure 4. Stars of all metallicity bins host warm super-Earths and warm sub-Neptunes. Cool Jupiters are intrinsically rare at all metallicities, but they are present in all four metallicity bins. We observe a steady increase of hot Jupiters with increasing metallicity. Sub-Saturns of all orbital periods are almost completely absent in the lower two metallicity bins, which represent half of the parent sample $\mathcal{S}$; they are almost exclusively found around high metallicity stars. While there are some examples of hot super-Earths in each metallicity bin, their numbers increase with increasing metallicity. Finally, there is almost a complete absence of hot sub-Neptunes in the lowest metallicity bins. Hot sub-Neptunes are more common with increasing metallicity.

In order to quantitatively assess the extent to which metallicity enhances the production of different types of planets, we compare the distribution of planet host metallicities to that of the field star population. In Table 5, we list the mean metallicities of the various planet subclasses as well as the standard error of the mean (SEM), which we compare to the mean field star metallicity, as measured from LAMOST spectra.

We assess the significance of the difference between field star and planet host star metallicities using the student t-test, which evaluates the difference between the means of the two samples in units of SEM, the t-statistic. The t-test also returns a p-value, which is the probability that field stars and planet host stars are drawn from distributions with the same mean value.

We may observe statistically significant differences in mean metallicity for two reasons: (1) intrinsic differences between planet host and field star metallicities or (2) residual offsets in the CKS and LAMOST metallicity scales. While we calibrated LAMOST metallicities to the CKS scale, we estimate that offsets of 0.01 dex may remain (see Section 2.4 and Appendix A). To account for the possibility of such residual offsets, we perform three t-tests for each sample where we shift the LAMOST metallicities by -0.01, 0.00, and $+0.01$ dex. Each of these different tests returns a different p-value. We use the largest (most conservative) p-value to assess differences in mean metallicities. If the largest p-value is less than 0.01, we deem the metallicity difference to be significant.

Stars hosting Jupiter-size planets have enhanced metallicities, $\langle[\mathrm{Fe/H}]\rangle = +0.12 \pm 0.04$ dex. The hot Jupiters hosts, with mean metallicity of $\langle[\mathrm{Fe/H}]\rangle = +0.19 \pm 0.04$ dex, are significantly enhanced relative to field stars (p-value $< 8 \times 10^{-4}$). While the mean metallicities for the warm and cool Jupiter hosts are also enhanced, the small numbers of such planets prevent a high significance detection of a metallicity enhancement.

Of all the planet size classes studied, the sub-Saturn hosts have the highest mean metallicity, $\langle[\mathrm{Fe/H}]\rangle = +0.16 \pm 0.02$ dex. Among the sub-Saturns, the hot sub-Saturns have the highest mean host star metallicity, $\langle[\mathrm{Fe/H}]\rangle = +0.26 \pm 0.04$ dex. We find that the hot and warm sub-Saturns hosts were significantly enhanced compared to field stars, while small numbers of detected cool sub-Saturns prevented a detailed comparison.

As a whole, the sub-Neptunes have a mean metallicity of $\langle[\mathrm{Fe/H}]\rangle = +0.05 \pm 0.01$ dex, close to the field star value. However, when split according to orbital period, we find that the hot sub-Neptunes have an enhanced mean metallicity,

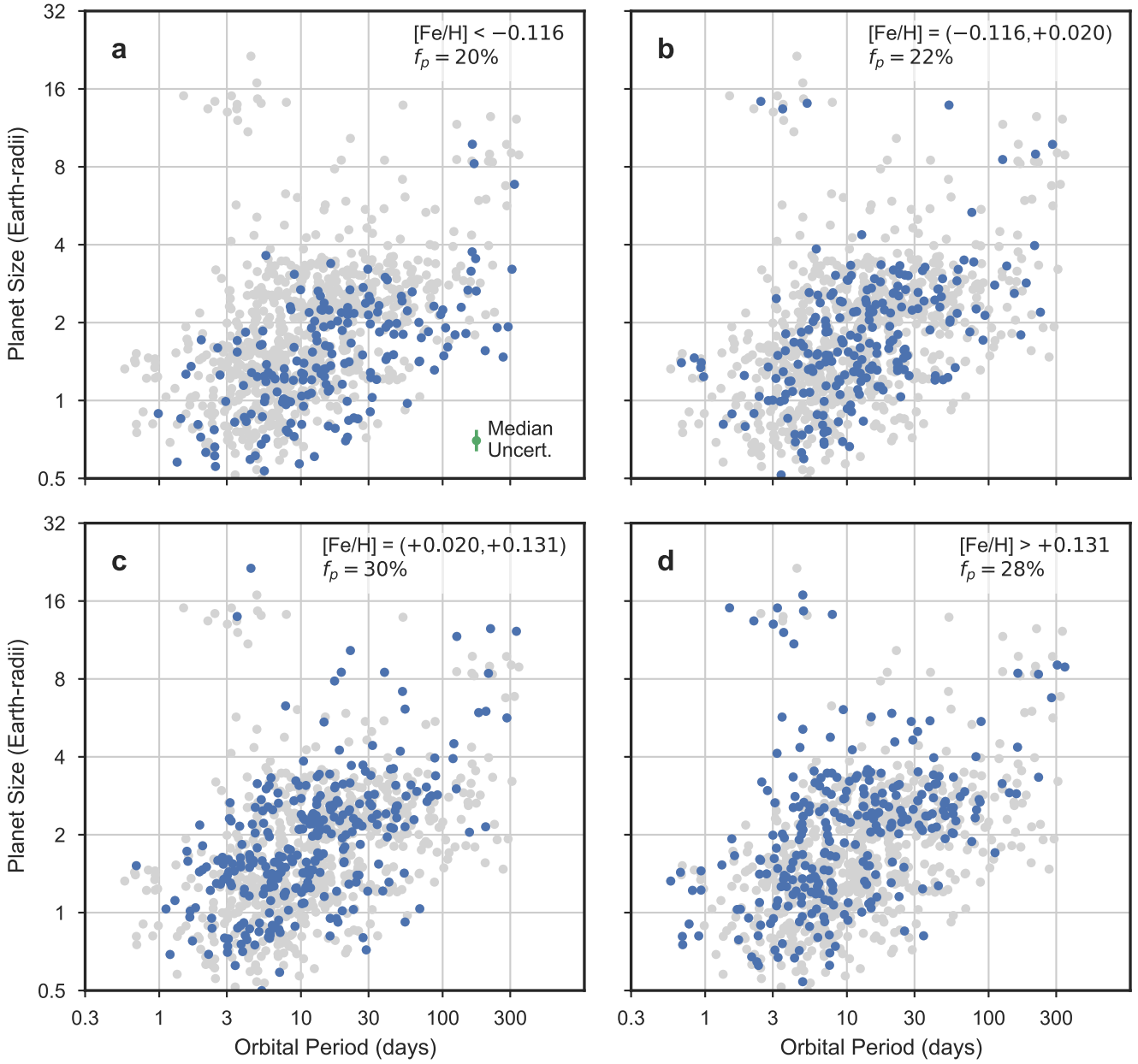

Figure 4. Orbital periods and radii of planets orbiting host stars belonging to different metallicity bins. Each metallicity bin captures an equal fraction (25%) of the parent stellar sample $\mathcal{S}$ (see Table 4). In panel (a), blue points represent planets orbiting host stars with [Fe/H] < -0.116 dex, the lowest metallicity bin. The gray points show the full planet sample $\mathcal{P}$ for context. At top right, we show f_p, the fraction of planets belonging to this metallicity bin. Panels (b)–(d): same as (a) except for different metallicity bins. If planets formed with equal efficiency regardless of host star metallicity, each bin would have $f_p = 25\%$ and the distribution of blue points would be indistinguishable from bin to bin. Planets around the lowest metallicity stars (a) are clearly confined to a restricted range of P–R_P space compared to planets around the highest metallicity hosts (d), notably in the lower right envelope of longer periods and smaller sizes.

$\langle$[Fe/H]$\rangle = +0.11 \pm 0.02$ dex. The large mean metallicity, combined with the large number of such planets ($N = 108$), results in a very significant detection of a metallicity enhancement (p-value $< 4 \times 10^{-9}$). Due to their high detectability and high intrinsic occurrence, the warm sub-Neptunes have the largest total number of any P–R_P subclass, $N = 282$. The large number of planets means that even small offsets in mean metallicities can be significant. While the mean metallicity of warm sub-Neptunes, $\langle$[Fe/H]$\rangle = +0.04 \pm 0.01$ dex), was not as high as that of the hot sub-Neptunes, it was significantly elevated relative to field stars (p-value $< 2 \times 10^{-3}$).

Finally, the super-Earth-size planets have the lowest mean metallicity of $+0.01 \pm 0.01$ dex, which is consistent with field star metallicity. However, as with the sub-Neptunes, the hot super-Earth hosts exhibit enhanced metallicity of $+0.05 \pm 0.01$ dex. The difference in mean metallicity (in dex) is not as large for the sub-Neptunes, so while the offset is still significant (p-value $< 2 \times 10^{-4}$), it is not as significant as for the hot sub-Neptunes. Despite the large number of warm super-Earths, we cannot detect a significant offset in mean metallicity.

One may wonder whether the mean metallicities of stars hosting warm super-Earths/sub-Neptunes could possibly be different from the mean metallicity of the *Kepler* field (-0.01 dex), given the high intrinsic occurrence of these planets. While numerous previous works have shown that warm super-Earths and sub-Neptunes are intrinsically common

Table 5
Comparison of Planet Host and Field Star Metallicities

Name	P day	R_P $R_\oplus$	$n_\star$	$\langle$[Fe/H]$\rangle$ dex	t-stat	p-value	Sig[a]
Hot Jupiters	1–10	8.0–24.0	14	$+0.19 \pm 0.04$	4.5	$<8 \times 10^{-4}$	Y
Warm Jupiters	10–100	8.0–24.0	4	$+0.06 \pm 0.02$	3.4	$<6 \times 10^{-2}$	N
Cool Jupiters	100–350	8.0–24.0	13	$+0.06 \pm 0.05$	1.4	$<3 \times 10^{-1}$	N
Hot Sub-Saturns	1–10	4.0–8.0	7	$+0.26 \pm 0.04$	6.5	$<7 \times 10^{-4}$	Y
Warm Sub-Saturns	10–100	4.0–8.0	19	$+0.14 \pm 0.02$	5.6	$<6 \times 10^{-5}$	Y
Cool Sub-Saturns	100–350	4.0–8.0	7	$+0.11 \pm 0.06$	1.9	$<1 \times 10^{-1}$	N
Hot Sub-Neptunes	1–10	1.7–4.0	108	$+0.11 \pm 0.02$	7.0	$<4 \times 10^{-9}$	Y
Warm Sub-Neptunes	10–100	1.7–4.0	282	$+0.04 \pm 0.01$	4.0	$<2 \times 10^{-3}$	Y
Cool Sub-Neptunes	100–350	1.7–4.0	29	-0.05 ± 0.03	-1.3	$<3 \times 10^{-1}$	N
Hot Super-Earths	1–10	1.0–1.7	181	$+0.05 \pm 0.01$	4.7	$<2 \times 10^{-4}$	Y
Warm Super-Earths	10–100	1.0–1.7	132	-0.04 ± 0.02	-2.2	$<1 \times 10^{-1}$	N
Cool Super-Earths	100–350	1.0–1.7	4	-0.36 ± 0.03	-11.4	$<2 \times 10^{-3}$	Y
All Jupiters	1–350	8.0–24.0	31	$+0.12 \pm 0.03$	4.2	$<6 \times 10^{-4}$	Y
All Sub-Saturns	1–350	4.0–8.0	33	$+0.16 \pm 0.02$	7.2	$<1 \times 10^{-7}$	Y
All Sub-Neptunes	1–350	1.7–4.0	419	$+0.05 \pm 0.01$	6.2	$<7 \times 10^{-7}$	Y
All Super-Earths	1–350	1.0–1.7	317	$+0.01 \pm 0.01$	1.3	$<8 \times 10^{-1}$	N

Note. We inspected the metallicity distribution of various groups of planets defined by their sizes and orbital periods. For each group, we computed the mean metallicity $\langle$[Fe/H]$\rangle$ and the standard error of the mean. We compare these metallicities to the metallicities of *Kepler* field stars, as measured by LAMOST, using the student t-test. This test returns a t-statistic, which quantifies the statistical difference in mean metallicities, and a p-value, the probability that the two samples were drawn from distributions with the same mean.

[a] Is the difference between planet host and field star metallicities significant (i.e., is p-value $< 10^{-2}$)?

(see, e.g., Petigura 2013), neither class of planets is found around 100% of stars. For example, in Section 5 we show that there are ≈17 warm super-Earths per 100 stars. Therefore, the mean metallicity of warm super-Earth hosts could be significantly different than that of field stars. We consider the following limiting case: Suppose that the process that produces warm super-Earths has a step function dependence on metallicity. Among *Kepler* targets, 17% of stars have [Fe/H] > 0.16 dex. A universe where every star with [Fe/H] > 0.16 dex produced one warm super-Earth, and every star with [Fe/H] < 0.16 dex produced zero, would be consistent with the measured occurrence. By consulting the distribution of *Kepler* field star metallicities (see Section 2.4), we find that the average metallicity of all stars with [Fe/H] > 0.16 dex is 0.25 dex. In this limiting case, the average metallicity of warm super-Earth hosts would be $\langle$[Fe/H]$\rangle = 0.25$ dex, much higher than the observed value of -0.04 ± 0.02 dex. That warm super-Earth and sub-Neptune hosts have mean metallicities nearly the same as that of field stars demonstrates qualitatively that their occurrence is not a steep function of metallicity. We show this quantitatively in Section 6.

In summary, close-in ($P < 10$ days) planets of all sizes have enhanced metallicity host stars. Warm super-Earth-size planets at intermediate distances ($P = 10$–100 days) orbit stars with a metallicity distribution similar to that of field stars. Larger planets in this period range have enhanced metallicities, and the sub-Neptunes and sub-Saturns have a significant metallicity enhancement. Finally, the number of detected cool planets ($P = 100$–350 days) of all sizes is too small to permit the detection of different host star metallicities.

4. Planet Occurrence: Methodology

In Section 3, we examined the number distribution of planets belonging to host stars of differing metallicities. We now address the underlying prevalence of planets. In this section, we describe our methodology for computing planet occurrence and fitting parameterized descriptions of planet occurrence to the observed data. Section 5 presents planet occurrence as a function of orbital period and planet size. Section 6 presents planet occurrence as a function of orbital period, planet size, and stellar metallicity.

4.1. Definitions

Adopting the notation of Youdin (2011), the probability that a star with properties z has a planet with properties $\boldsymbol{x}$ that lie in a volume of $d\boldsymbol{x}$ of phase space is

$$df = \frac{\partial f(\boldsymbol{x}, z)}{\partial \boldsymbol{x}} d\boldsymbol{x}. \quad (1)$$

In this paper, z is metallicity $M = $ [Fe/H],[14] and $\boldsymbol{x}$ is some combination of $\log P$ and $\log R_P$.[15] As a matter of convenience we express differential distribution as

$$\frac{\partial f(\boldsymbol{x}, z)}{\partial \boldsymbol{x}} = C g(\boldsymbol{x}, z), \quad (2)$$

where C is a normalization constant and g is a shape function that may depend on stellar and/or planetary properties.

The number of planets per star (NPPS) having stellar properties z over a specified volume of planet properties X is

$$f(z) = \int_X C g(\boldsymbol{x}, z) d\boldsymbol{x}. \quad (3)$$

The NPPS having stellar properties within a range of stellar properties Z is

$$f = \frac{1}{n_\star} \sum_j \int_X C g(\boldsymbol{x}, z_j) d\boldsymbol{x}, \quad (4)$$

where j labels stars where $z_j \in Z$.

[14] We use M interchangeably with [Fe/H] for compactness.

[15] In this paper, log only refers to $\log_{10}$. Natural logs are always expressed as ln.

Note that our definition of planet occurrence is NPPS. As a matter of convenience, we often express this in units of planets per 100 stars. Some papers (e.g., Fischer & Valenti 2005) define planet occurrence as the fraction of stars with planets (FSWP). For a transit survey like *Kepler*, computing FSWP is challenging because one must compute the number of stars *without planets*. However, a transit survey cannot distinguish between stars *without planets* and stars *without transiting planets*. Converting between NPPS and FSWP requires detailed modeling of planet multiplicity and mutual inclinations and is beyond the scope of this work. For more discussion on these points, see Youdin (2011) and references therein.

4.2. Occurrence within a Cell

In this paper, we compute and display planet occurrence over rectilinear "cells" constructed from various combinations of $\log P$, $\log R_P$, and M. We express the size of these cells as $\Delta \log P$, $\Delta \log R_P$, ΔM, measured in dex. The occurrence within a cell is $f_{\rm cell}$, which depends on the number of independent trials $n_{\rm trial}$ that could have yielded a detected planet. In the limit where every star in $\mathcal{S}$ has one transiting planet that is also detectable by the *Kepler* pipeline, $n_{\rm trial} = n_\star$. In practice, $n_{\rm trial}$ must account for (1) non-transiting orbital inclinations and (2) lack of detectability due to insufficient S/N. For a given planet size and orbital period,

$$n_{\rm trial} = \sum_{i=1}^{n_\star} p_{{\rm tr},i}(P) p_{{\rm det},i}(P, R_P), \tag{5}$$

where i labels each star in $\mathcal{S}$, $p_{{\rm tr},i}$ is the transit probability, and $p_{{\rm det},i}$ is the probability that a transiting planet would be detectable by the *Kepler* pipeline. We represent this more compactly as

$$n_{\rm trial} = n_\star \langle p_{{\rm tr},i}\, p_{{\rm det},i} \rangle, \tag{6}$$

where $\langle \cdot \rangle$ denotes the arithmetic mean.

Following Bowler et al. (2015), if we assume that planet occurrence is log-uniform within a cell, $n_{\rm trial}$ for a cell of size $\Delta \boldsymbol{x} = \Delta \log P \times \Delta \log R_P$ is

$$n_{\rm trial} = \frac{n_\star}{\Delta \boldsymbol{x}} \int \langle p_{\rm tr} p_{\rm det} \rangle d\boldsymbol{x}. \tag{7}$$

For a given cell with $n_{\rm trial}$ trials, $n_{\rm pl}$ detected planets, and $n_{\rm nd} = n_{\rm trial} - n_{\rm pl}$ non-detections, the likelihood of $f_{\rm cell}$ is given by the binomial distribution

$$P(f_{\rm cell}|n_{\rm pl}, n_{\rm trial}) = P(n_{\rm pl}|f_{\rm cell}, n_{\rm trial}) \tag{8}$$

$$= C f_{\rm cell}^{n_{\rm pl}} (1 - f_{\rm cell})^{n_{\rm nd}}, \tag{9}$$

where C is the following normalization constant:[16]

$$C = \frac{(n_{\rm trial} + 1)\Gamma(n_{\rm trial} + 1)}{\Gamma(n_{\rm pl} + 1)\Gamma(n_{\rm nd} + 1)}. \tag{10}$$

If no transits are detected in a given cell, we may place an upper bound on the occurrence rate by calculating the maximum value of $f_{\rm cell}$ that yields zero planet detections ($n_{\rm pl} = 0$) with 90% probability. This is done by numerically finding an $f_{\rm cell}$ that satisfies

$$\int_0^{f_{\rm cell}} P(f|n_{\rm pl}, 0) df = 90\%. \tag{11}$$

When analyzing planet occurrence as a function of metallicity, we must consider the fraction of stars in $\mathcal{S}$ that falls within a specified range of metallicity. Therefore, to compute $f_{\rm cell}$ for a cell bounded by $[R_{P,1}, R_{P,2}]$, $[P_1, P_2]$, and $[M_1, M_2]$, we simply multiply $n_{\rm trial}$ from Equation (7) by the fraction of LAMOST stars with metallicities between M_1 and M_2. This treatment assumes that planet detectability does not depend on metallicity. We justify this assumption in Appendix B.

4.3. Completeness

To compute occurrence, we must estimate $\langle p_{\rm tr} p_{\rm det} \rangle$. The probability that a randomly inclined planet at a distance of a will transit with $b < 0.9$ is $p_{\rm tr} = 0.9\ R_\star/a$, assuming circular orbits. We compute $p_{\rm tr}$ for each star using the tabulated values of $R_\star$ and $M_\star$ from the Q1–Q17 (DR25) stellar properties table (Mathur et al. 2017), along with Kepler's Third Law.

The probability of detecting a transiting planet with size R_P and period P depends on the number of factors, the most significant of which is the S/N. We follow the methodology of Fulton et al. (2017) and compute an expected S/N using the following formula:

$$\mathrm{S/N} = \left(\frac{R_P}{R_\star}\right)^2 \sqrt{\frac{T_{\rm obs}}{P}} \left(\frac{1}{\sigma(T_{14})}\right). \tag{12}$$

Here, $T_{\rm obs}$ is the time the star was observed by *Kepler*, and $\sigma(T_{14})$ represents the photometric variability on transit timescales. The DR25 stellar properties table lists $T_{\rm obs}$, $R_\star$, and photometric noise[17] on 3, 6, and 12 hr timescales. We compute $\sigma(T_{14})$ at intermediate values of T_{14} through piecewise linear interpolation (or extrapolation) in log σ–logT_{14} space.

Characterizing $p_{\rm det}$ as a function of S/N has been addressed in several previous works (e.g., Fressin et al. 2013; Petigura et al. 2013; Christiansen et al. 2015). This is a challenging problem, which depends on how a complex, multi-stage transit pipeline performs in the face of correlated, non-stationary, and non-Gaussian photometric noise. We adopted the $p_{\rm det}$ (S/N) of Fulton et al. (2017), who used the transit injections of Christiansen et al. (2015) to derive the following relationship:

$$p_{\rm det}(s) = \Gamma(k) \int_0^{\frac{s-l}{\theta}} t^{k-1} e^{-t} dt, \tag{13}$$

where s is the S/N, $k = 17.56$, $l = 1.00$, and $\theta = 0.49$. We illustrate $\langle p_{\rm det} \rangle$ and $\langle p_{\rm tr} p_{\rm det} \rangle$ as a function of P and R_P in Figure 5.

4.4. Fitting the Occurrence Distribution

We often wish to characterize the occurrence distribution with parametric models. These models serve to quantify shapes and trends in the population of planets. We extend the maximum-likelihood method of Howard et al. (2012) to fit $Cg(\boldsymbol{x}, \boldsymbol{z}; \boldsymbol{\theta})$ to the observed planet population, where $\boldsymbol{\theta}$ represents the vector of shape parameters in our model.

[16] Here, the factorials have been replaced with gamma functions via $\Gamma(x + 1) = x!$, and $(n + 1)$ is a normalization constant.

[17] Combined Differential Photometric Precision (CDPP; Jenkins et al. 2010).

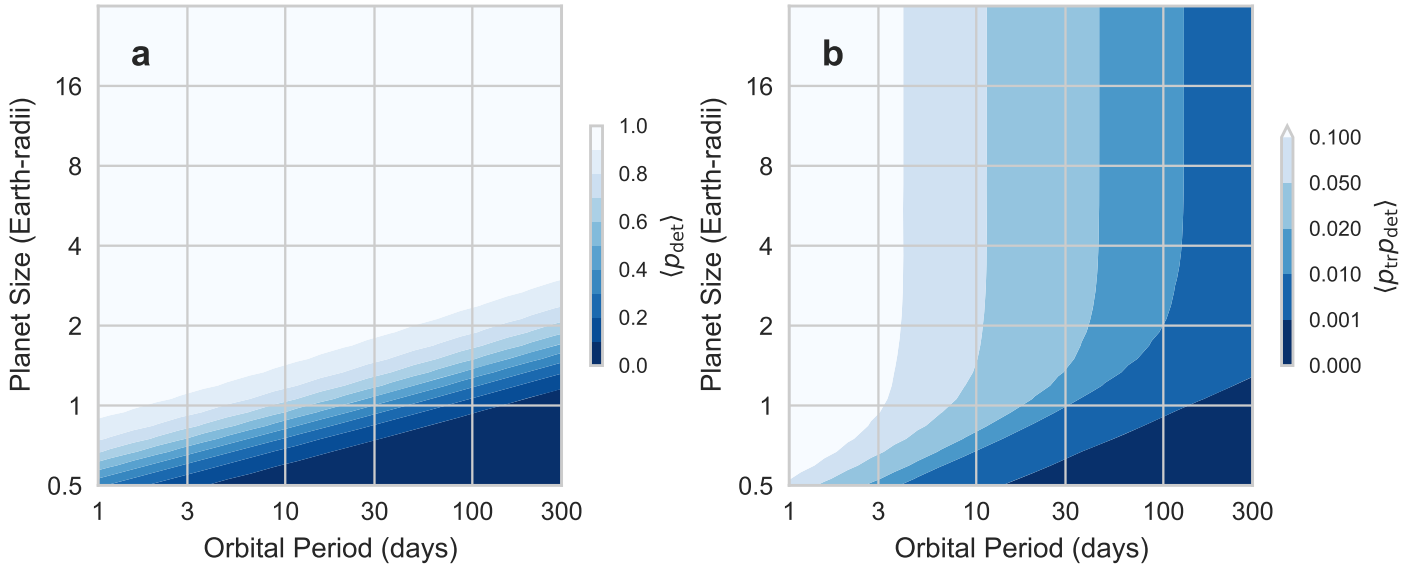

Figure 5. Average detectability of planets as a function of planet size and orbital period. Panel (a): the probability that a transiting planet would be detected averaged over all stars in our sample, $\langle p_{\rm det} \rangle$. Panel (b): the product of the transit and detection probability averaged over all stars in the sample $\langle p_{\rm det}\, p_{\rm tr} \rangle$. The effective number of stars from which the planet detections are drawn is simply $n_{\rm trial} = n_\star \langle p_{\rm det}\, p_{\rm tr} \rangle$.

The occurrence within each cell $f_{\rm cell}$ gives an estimate of the differential occurrence rate via

$$Cg(\boldsymbol{x}, z; \boldsymbol{\theta}) = \frac{f_{\rm cell}}{\Delta \boldsymbol{x}}. \tag{14}$$

The log-likelihood of a given model with parameters $\{C, \boldsymbol{\theta}\}$ given the observed rate in a single cell labeled by i is

$$\ln L_i = n_{{\rm pl},i} \ln Cg\Delta \boldsymbol{x} + n_{{\rm nd},i} \ln(1 - Cg\Delta \boldsymbol{x}). \tag{15}$$

Each cell is an independent constraint on $Cg(\boldsymbol{x}, z; \boldsymbol{\theta})$, so a fit to multiple cells requires maximizing the combined log-likelihood over all cells:

$$\ln L = \sum_{i=1}^{n_{\rm cell}} L_i. \tag{16}$$

When fitting parameterized models of occurrence in Sections 5 and 6, we use very small bins having $\Delta \log P = 0.05$ dex and $\Delta \log M = 0.05$ dex. As a result, many bins have one or zero planet detections. The maximum-likelihood method is stable for small bin sizes because Equation (15) incorporates non-detections. We experimented with even finer bins, but did not observe significant changes in the results. After finding the max-likelihood model, we explore the credible range of models using Markov Chain Monte Carlo (MCMC) sampling.[18]

5. Planet Occurrence: Period and Radius

We first present planet occurrence on the P–R_P plane, averaging over stellar metallicity. Following Howard et al. (2012), we divided the domain of orbital period and planet size into numerous cells spanning $\Delta \log P \times \Delta \log R_P = 0.25$ dex $\times$ 0.15 dex. For each cell, we computed $f_{\rm cell}$ according to the prescription from Section 4.2. We display $f_{\rm cell}$ as a checkerboard in Figure 6, where each cell in is shaded and annotated according to the occurrence of planets within the cell. We list the cell-by-cell rates, uncertainties, and upper limits in Table 8 in the Appendix.

We also show occurrence as a function of orbital period for various size classes in Figure 7, computed over bins spanning $\Delta \log P = 0.25$ dex. Errorbars and upper limits are computed according to the methodology presented in Section 4.2. For super-Earths and sub-Neptunes, inspection of the binned occurrence rates reveals that $df/d\log P$ increases with P until a transition period P_0, above which occurrence is nearly uniform in $\log P$. This "knee" points toward an important orbital distance in the formation or subsequent migration of planets. We characterized the period distribution with the following parameterization from Howard et al. (2012):

$$\frac{df}{d\log P} = CP^\beta (1 - e^{-(P/P_0)^\gamma}). \tag{17}$$

Far from P_0, this function reverts to a power law with the following indices:

$$\frac{df}{d\log P} \propto \begin{cases} P^\alpha & \text{if } P \ll P_0, \text{ where } \alpha = \gamma + \beta \\ P^\beta & \text{if } P \gg P_0. \end{cases} \tag{18}$$

We fit this distribution using the methodology from Section 4.2 and list the associated parameters in Table 6. Our best-fit model and 1σ range of credible models are shown in Figure 7. As explained in Section 4.4, the max-likelihood fitting uses much finer bins than those displayed in Figure 7. The binned rates shown in Figure 7 serve to guide the eye and are not used in the fitting.

For super-Earths the transition period P_0 is $6.5^{+1.6}_{-1.2}$ days. At shorter orbital periods, occurrence rises rapidly with increasing P, with $df \propto P^\alpha d\log P$, where $\alpha = 2.4^{+0.4}_{-0.3}$. At longer orbital periods, $df \propto P^\beta d\log P$, where $\beta = -0.3^{+0.2}_{-0.2}$. The transition period for sub-Neptunes is farther out at $P_0 = 11.9^{+1.7}_{-1.5}$ days, but the power law indices are similar, with $\alpha = 2.3^{+0.2}_{-0.2}$ and $\beta = -0.1^{+0.1}_{-0.1}$. The distributions of sub-Saturns and Jupiters are

[18] We used the affine invariant sampler of Goodman & Weare (2010), as implemented in Python by Foreman-Mackey et al. (2013).

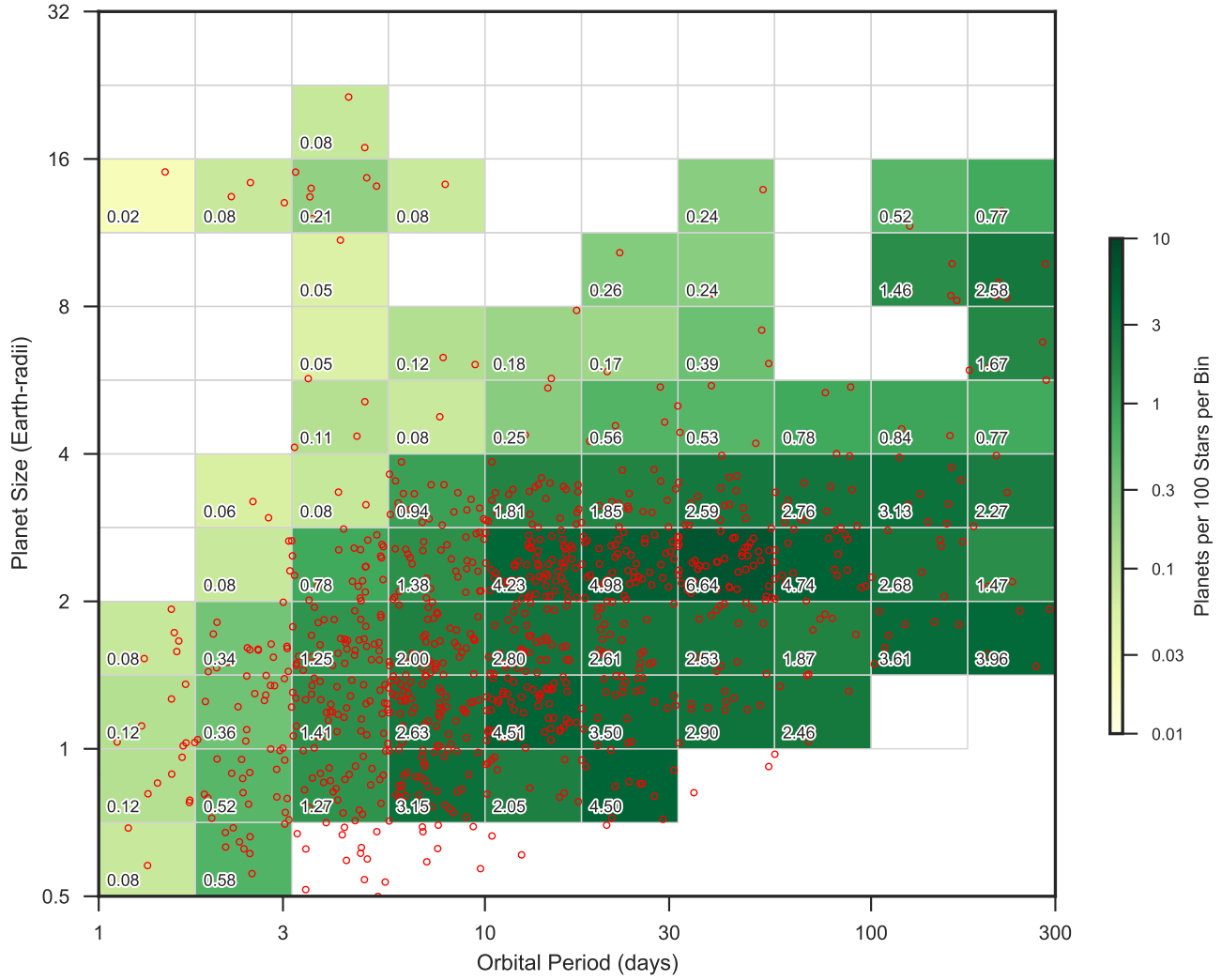

Figure 6. The domain of orbital period and planet size, which has been sub-divided into numerous sub-domains, "cells." Each cell is annotated with the average number of planets per 100 stars having properties within each cell. This quantity is also reflected through the shading of each cell. The number of planets has been corrected for the probability of transiting and for detection completeness.

not well-described by the power law cutoff model. Their occurrence gradually increases over $P = 1$–300 days.

Figures 6 and 7 are convenient tools for calculating planet occurrence over various domains of period and radius. However, several features in the planet population are hard to see in these plots due to the coarse bin sizes and the arbitrary location of the bin boundaries. Figure 8 shows a finer view of planet occurrence. We computed occurrence in bins of period and radius spanning 0.25 dex and 0.10 dex, respectively. We then shifted the bins in small steps of $\Delta \log P$ and $\Delta \log R_P$ to smoothly trace out planet occurrence in the P–R_P plane. The shading in Figure 8 gives a bird's eye view of the prevalence of various types of planets.

Panels (a) and (b) of Figure 8 show occurrence on a linear scale to highlight the most abundant types of planets (super-Earths and sub-Neptunes). We note the rapid increase in occurrence for $R_P \lesssim 4\ R_\oplus$, which has been noted by Howard et al. (2012) and numerous subsequent works. Recently, Fulton et al. (2017) re-examined the radius distribution of small planets using the CKS catalog. Fulton et al. (2017) also noted a gap in the radius distribution at $R_P = 1.7\ R_\oplus$. This gap was predicted by various groups who modeled the erosion erosion of envelopes (e.g., Lopez & Fortney 2013; Owen & Wu 2013; Jin et al. 2014; Chen & Rogers 2016). We also resolve the gap that separates the super-Earth and sup-Neptune populations.

Panels (c) and (d) of Figure 8 show occurrence on a logarithmic scale to highlight domains of low planet occurrence. Although hot Jupiters are rare, they constitute a distinct island in the P–R_P plane, surrounded by a sea of still lower occurrence. Hot planets of intermediate sizes are very rare. This triangular "Hot Planet Desert" has been noted by Mazeh et al. (2016) and Dong et al. (2018), and is thought to be due to photo-evaporative envelope stripping.

We also observe an factor of ≈5–10 increase in occurrence of Jupiters from $P = 100$–300 days. Cumming et al. (2008) analyzed the planets detected by radial velocities from the Keck Planet Search and observed that giant planet occurrence increases by a factor ≈5 from $P = 100$–300 days. This rise in the occurrence of Jovian planets between 100–300 days, observed in both the *Kepler* stellar population and among nearby stars, may be associated with an ice-line at ∼1 au.

The joint distribution of planets in the P–R_P plane shown in Figure 8 is an important quantitative description of the population of planets. For example, this distribution is useful for studies predicting planet yields in future surveys (e.g., Sullivan et al. 2015). To facilitate such studies, we sample this

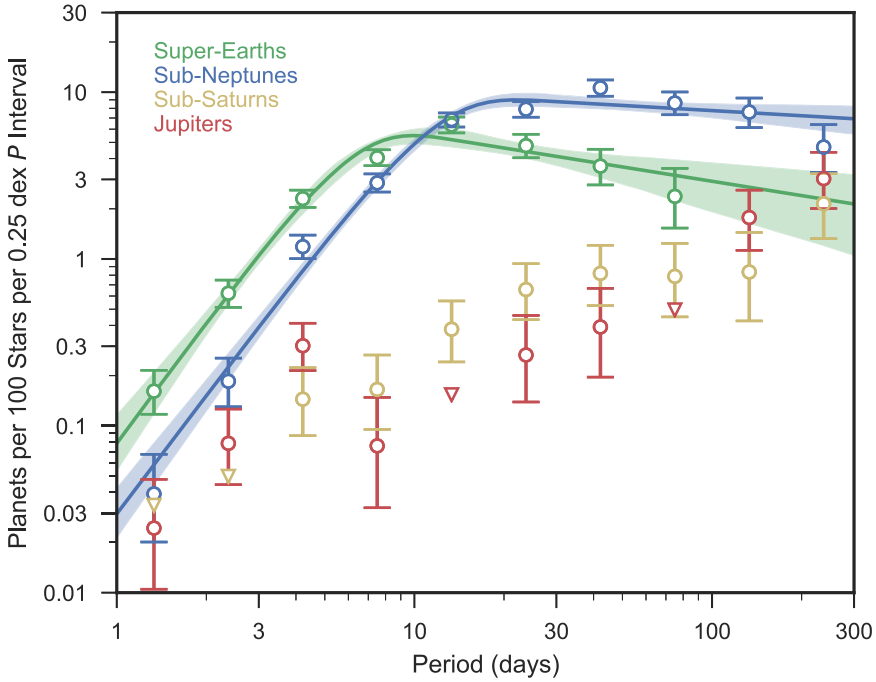

Figure 7. Planet occurrence as a function of orbital period for various size classes. For example, green points show the number of super-Earths per 100 stars per 0.25 dex interval in the period. Downward arrows represent upper limits (90%). For the super-Earths and sub-Neptunes, we fit the occurrence using the power law and exponential cutoff model described in Section 5 (Equation (17)). The solid lines and bands show the best-fitting model and 1σ range of credible models, respectively. For super-Earths and sub-Neptunes at $P < 10$ days, planet occurrence increases like $df/d\log P \propto P^{\alpha}$, where $\alpha \approx 2.4^{+0.4}_{-0.3}$ and $\alpha \approx 2.3^{+0.2}_{-0.2}$, respectively. At longer orbital periods, occurrence is nearly uniform in $\log P$. A transition period P_0 characterizes where the distributions changes slope, which occurs at $P_0 = 6.5^{+1.6}_{-1.2}$ days and $P_0 = 11.9^{+1.7}_{-1.5}$ days, respectively. The distributions of sub-Saturns and Jupiters are not well-described by this model. Their occurrence gradually increases over $P = 1$–300 days.

distribution and provide the periods and radii for a representative population of planets in a sample of 100,000 Sun-like stars in Table 9. This synthetic population also offers an convenient way to compute integrated planet occurrence (but not uncertainties) over arbitrary bins of P and R_P.

6. Planet Occurrence: Period, Radius, and Stellar Metallicity

Here, we analyze planet occurrence as a function of period, radius, and stellar metallicity. As a first step, we divided the super-Earth and sub-Neptune populations into two groups, depending on whether their host stars had sub- or super-solar metallicities. We then repeated the analysis from Section 5, modeling the occurrence distribution of both groups according to Equation (17).

Figure 9 shows the occurrence of super-Earths and sub-Neptunes as a function of orbital period for sub- and super-solar metallicity hosts. The occurrence rates for $P > 10$ days are generally consistent for the two metallicity bins. For $P < 10$ days, the occurrence rates of super-Earths/sub-Neptunes is ≈2–3 times higher for super-solar metallicity hosts compared to sub-solar metallicity hosts.

To quantify the metallicity correlation for our different planet classes, we modeled occurrence using the following parametric function:

$$\frac{df}{d\log P\, dM} = CP^{\alpha}10^{\beta M}. \tag{19}$$

This model extends the following exponential model,

$$\frac{df}{dM} = C10^{\beta M}, \tag{20}$$

which was used in Fischer & Valenti (2005) and in several subsequent works. Given that $M = \log(n_{\rm Fe}/n_{\rm H}) - \log(n_{\rm Fe}/n_{\rm H})_{\odot}$, Equation (20) is equivalent to a power law relationship between planet occurrence and the number density of iron atoms in a star's photosphere relative to hydrogen,

$$\frac{df}{dM} = C\left[\frac{n_{\rm Fe}/n_{\rm H}}{(n_{\rm Fe}/n_{\rm H})_{\odot}}\right]^{\beta}. \tag{21}$$

We fit the distribution defined in Equation (19) using the methodology from Section 4.4 and list the associated model parameters in Table 7.

For the hot super-Earths we found $\beta = +0.6^{+0.2}_{-0.2}$, indicating a significant positive metallicity correlation. That $\beta \sim 1$ for hot super-Earths means that their occurrence is nearly proportional to the number of iron atoms in a star's photosphere, relative to hydrogen. Through our MCMC modeling, we found that the metallicity index β was only weakly correlated with the period index α. In contrast, α and the normalization constant C were highly covariant.

The metallicity correlation steepens for larger planets. For hot sub-Neptunes, $\beta = +1.6^{+0.3}_{-0.3}$. For hot sub-Saturns and hot Jupiters, our planet sample $\mathcal{P}$ contains only 7 and 14 planets, respectively. Small sample size leads to larger uncertainties on β, and for hot sub-Saturns and Jupiters, $\beta = +5.5^{+1.6}_{-1.5}$ and $\beta = +3.4^{+0.9}_{-0.8}$, respectively. Despite these larger

Table 6
Best-fit Parameters for Planet Period Distributions

Size Class	R_P	[Fe/H]	$\log_{10} C$	β	P_0	γ
Super-Earth	1.0–1.7	all	$-0.32^{+0.32}_{-0.25}$	$-0.3^{+0.2}_{-0.2}$	$6.5^{+1.6}_{-1.2}$	$2.7^{+0.3}_{-0.2}$
		<0	$-0.11^{+0.51}_{-0.39}$	$-0.4^{+0.3}_{-0.4}$	$8.8^{+2.3}_{-1.8}$	$3.0^{+0.4}_{-0.3}$
		>0	$-0.35^{+0.32}_{-0.23}$	$-0.4^{+0.2}_{-0.3}$	$5.1^{+1.4}_{-0.8}$	$2.9^{+0.4}_{-0.3}$
Sub-Neptune	1.7–4.0	all	$-0.28^{+0.18}_{-0.18}$	$-0.1^{+0.1}_{-0.1}$	$11.9^{+1.7}_{-1.5}$	$2.4^{+0.2}_{-0.2}$
		<0	$-0.75^{+0.27}_{-0.22}$	$+0.1^{+0.1}_{-0.2}$	$10.6^{+2.5}_{-1.8}$	$2.7^{+0.4}_{-0.4}$
		>0	$+0.09^{+0.29}_{-0.24}$	$-0.3^{+0.1}_{-0.2}$	$13.2^{+2.8}_{-2.1}$	$2.5^{+0.2}_{-0.2}$

Note. Best-fit parameters associated with the occurrence model defined in Equation (17).

uncertainties, it is clear that hot sub-Saturns and Jupiters have significantly steeper metallicity dependencies than hot super-Earths and sub-Neptunes.

Our occurrence rate model depends on both P and M, and thus cannot be displayed on a 2D plot. For display purposes, we performed an integration from $P = 1$–10 days. Figure 10 shows the number of planets belonging to different size classes per 100 stars having $P = 1$–10 days for various metallicity intervals spanning $\Delta M = 0.2$ dex. As in Figure 7, the binned rates serve to guide the eye and are not used in the fitting.

We observe different metallicity dependencies for more distant planets. As shown in Figure 10, warm super-Earths are consistent with no metallicity dependence $\beta = -0.3^{+0.2}_{-0.2}$. Warm sub-Neptunes show a positive metallicity correlation, but their power law index $\beta = +0.5^{+0.2}_{-0.2}$ is significantly shallower than that of the hot sub-Neptunes. The warm sub-Saturns have a positive metallicity correlation of $\beta = +2.1^{+0.7}_{-0.7}$, again more shallow than that of the hot sub-Saturns. Finally, our sample contains too few warm Jupiters (four) to search for a metallicity correlation.

In summary, some but not all planet subclasses are associated with stellar metallicity. Warm super-Earths exhibit no metallicity dependence. The planet-metallicity correlation steepens with increasing size and decreasing orbital period, with the hot Jupiters and hot sub-Saturns showing the steepest metallicity dependencies. We discuss some physical interpretation for these trends in the following section.

7. Summary and Discussion

The *Kepler* survey has revealed many valuable insights regarding the prevalence of planets with $P \lesssim 1$ year. While numerous previous works have addressed *Kepler* planet occurrence in the P–R_P plane,[19] we have leveraged the high-precision, high purity CKS catalog to produce a clearer picture of planet occurrence as a function of orbital period and size (Section 5; Figures 6–8). This complements the work of Fulton et al. (2017) that focused on super-Earth and sub-Neptune occurrence rates with $P = 1$–100 days. Our occurrence rates may be easily incorporated into yield calculations for future surveys sensitive to planets with $P \lesssim 1$ year. We found a hot Jupiter occurrence rate of $0.57^{+0.14}_{-0.12}$ %, which we compare to results from RV surveys in Section 7.3. While hot Jupiters are rare, we find that they occupy a distinct island in P–R_P space, surrounded by a sea of still lower occurrence.

[19] A non-exhaustive list includes Youdin (2011), Catanzarite & Shao (2011), Traub (2012), Howard et al. (2012), Fressin et al. (2013), Dong & Zhu (2013), Petigura et al. (2013), Foreman-Mackey et al. (2014), Burke et al. (2015), Dressing & Charbonneau (2015), and Mulders et al. (2015).

Using the CKS catalog combined with LAMOST metallicities for a large sample of *Kepler* field stars, we have computed planet occurrence as a function of host star metallicity (Section 6, Figures 9 and 10). We observe a clear association between stellar metallicity and the prevalence of certain types of planets. Figure 11 summarizes our findings: stellar metallicity is associated with the occurrence of hot planets ($P < 10$ days) and with the occurrence of large planets ($R_P > 1.7\,R_\oplus$, $P < 100$ days).

While the occurrence of some types of planets correlate strongly with metallicity, others exhibit only a moderate or negligible correlation. The interpretation of these metallicity trends is not yet clear. There are many ways in which differences in metallicity might alter the processes of planet formation and evolution: through differences in the surface density of solids in the protoplanetary disk, the growth rate and abundance of solid cores, the availability of gas, the efficiency of gas accretion onto protoplanets, or the rate of disk migration. There may also be others. We offer some interpretations of these trends in Sections 7.1–7.3. Finally, we consider future studies that could extend our understanding of the connection between planets and host star metallicity in Section 7.4.

7.1. Metallicity and the Prevalence of Large Planets

We found that metallicity is associated with the prevalence of large planets. This trend is consistent with the planet-metallicity correlation for Jovian-mass planets found by RV surveys (Santos et al. 2004; Fischer & Valenti 2005) and with previous analyses of the metallicities of *Kepler* planet hosts (e.g., Buchhave et al. 2012). We observed that for smaller planets, the metallicity correlation weakens, which is consistent with trends observed in previous RV studies (e.g., Sousa et al. 2008; Ghezzi et al. 2010) and metallicity studies of *Kepler* planet hosts (e.g., Buchhave et al. 2012, 2014). Our analysis builds upon these previous studies by providing planet occurrence as a function of metallicity based on the high-precision, high purity CKS catalog of planet radii and metallicities.

There are roughly 15 warm super-Earths per 100 stars over the metallicities ranging from −0.4 to +0.4 dex. If we assume that stellar metallicity traces disk solid surface densities from 0.10 to 0.40 au (which correspond $P \approx 10$–100 day orbits), then warm super-Earths can form with moderate efficiency, even in disks with low solid surface densities. Moreover, a factor of 6 increase in solids does not enhance the likelihood of a star to host a warm super-Earth.

RV and TTV mass measurements of super-Earths typically find masses of 1–10 $M_\oplus$ (Marcy et al. 2014, see Figure 7 of Sinukoff et al. 2017 for an updated census). The bulk densities of these planets are usually consistent with combinations of rock and iron, but are inconsistent with even small envelope

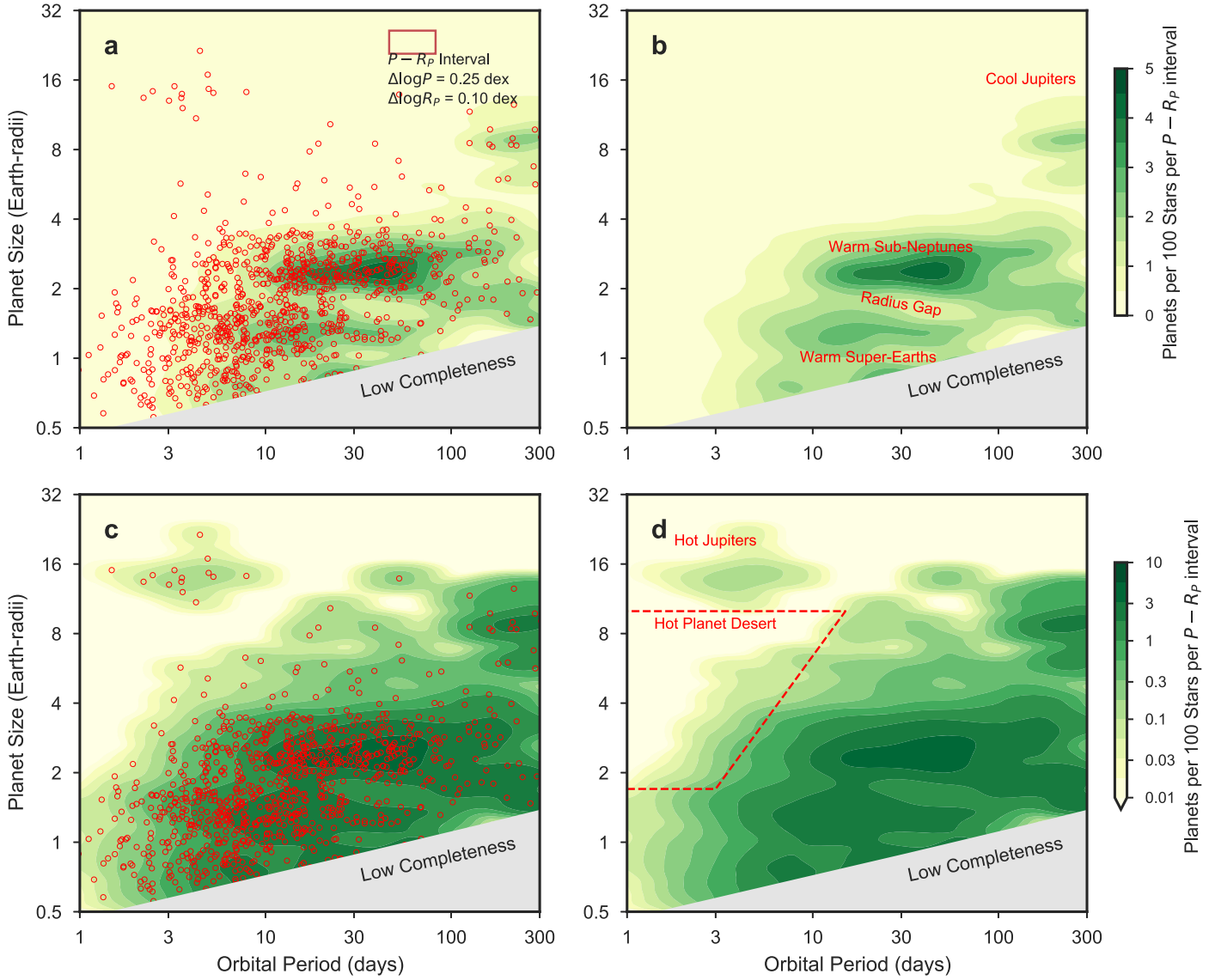

Figure 8. Panel (a) shows the planet sample $\mathcal{P}$ and planet occurrence in the P–R_P plane. The shading at each (P, R_P) point represents the number of planets per 100 stars within an interval centered at (P, R_P) that spans 0.25 dex in log P and 0.10 dex in log R_P. The size of the interval is indicated at top right. For example, the darkest contour indicates that there are ≈4 planets per 100 Sun-like stars having periods within 0.125 dex of 40 days and radii within 0.05 dex of 2.5 $R_\oplus$. Panel (b): same as (a), but without detected planets for clarity. Warm sub-Neptunes are the most abundant class of planets, and warm super-Earths are another common class of planets. The super-Earths and sub-Neptunes are separated by a diagonal gap of low planet occurrence, described by Fulton et al. (2017). Panels (c)–(d): same as (a)–(b), but with logarithmic shading to highlight domains of low occurrence. The hot Jupiter population is an "island," distinct from the rest of the planet population. There is a "desert" of hot planets having intermediate sizes.

fractions of $M_{\rm env}/M_P \sim 1\%$, which would significantly reduce the observed bulk densities.

Even though sub-Neptunes are significantly larger than super-Earths, they share several important characteristics. Sub-Neptunes are typically 5–15 $M_\oplus$, which overlaps with the super-Earth mass range. Sub-Neptune bulk densities require H/He envelopes that are 1%–10% of the planet's mass, which significantly enlarge the planets without significantly altering their masses. Therefore, the typical sub-Neptune contains a comparable amount of solid material as a typical super-Earth. Also, their overall occurrence rates are roughly comparable at 5%–10% per 0.25 dex period interval (Figure 9).

We see an important difference between super-Earths and sub-Neptunes when we compare their dependencies on stellar metallicity. There are comparable numbers of warm super-Earths and warm sub-Neptunes around stars having [Fe/H] = [−0.4, −0.2] dex, about 20 per 100 stars. Unlike the warm super-Earths, the occurrence of warm sub-Neptunes increases with metallicity. There are about 40 warm sub-Neptunes per 100 stars having [Fe/H] = [+0.2, +0.4] dex.

Perhaps the enriched disks of metal-rich stars form super-Earths more efficiently, but they easily acquire the few percent envelopes and are converted into sub-Neptunes. This could occur if super-Earths form more quickly in metal-rich disks, allowing for more time to accrete 1%–10% gaseous envelopes.

Such a timing argument was proposed by Dawson et al. (2015), to explain an apparent absence of warm super-Earths around the highest metallicity stars ([Fe/H] > 0.18 dex) in the Buchhave et al. (2014) metallicity catalog.[20] While only 12/132

[20] The Dawson et al. (2015) definition of warm super-Earths was slightly different than ours: $R_P < 1.5\ R_\oplus$, $P < 15$ days.

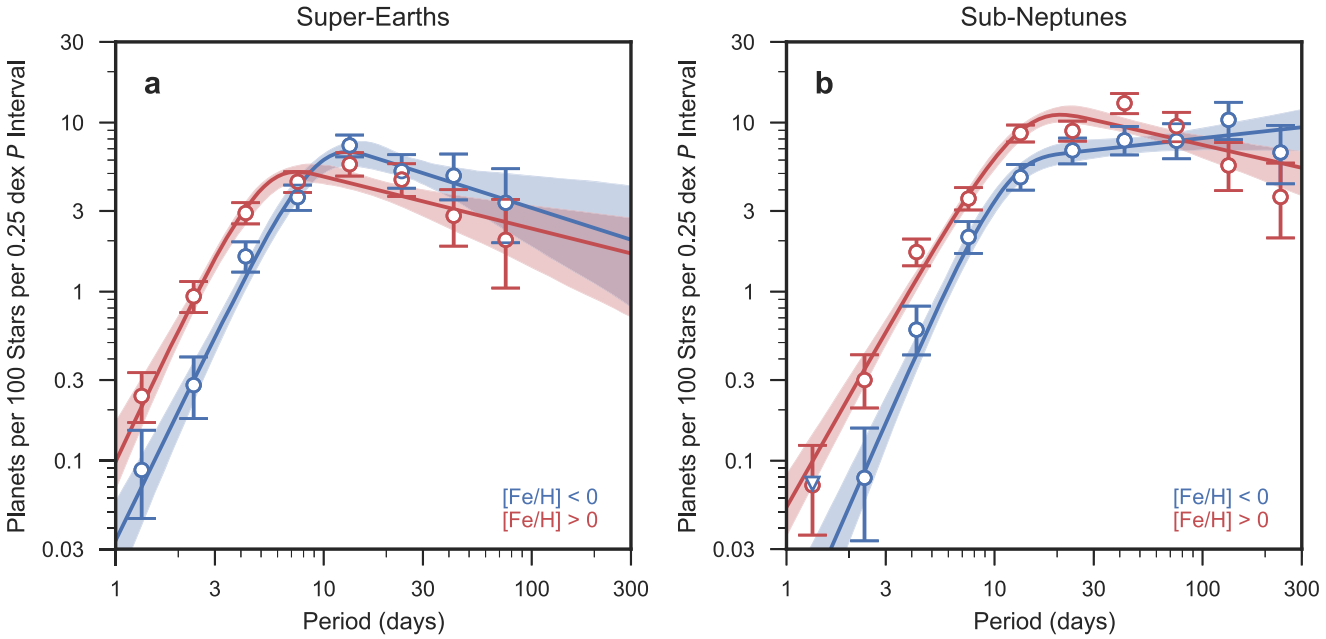

Figure 9. Panel (a) is analogous to Figure 7, except showing the period distribution of super-Earths at different host star metallicities. For example, red points show the number of super-Earths per 100 stars with super-solar metallicity per 0.25 dex period interval. Hot super-Earths are more common around metal-rich stars. Panel (b) same as (a) except for sub-Neptunes. Hot sub-Neptunes are also more common around metal-rich stars.

Table 7
Best-fit Parameters for Planet-metallicity Distributions

	$\log_{10} C$	α	β
Hot Super-Earth	$-2.22^{+0.09}_{-0.09}$	$+1.7^{+0.1}_{-0.1}$	$+0.6^{+0.2}_{-0.2}$
Warm Super-Earth	$+0.05^{+0.22}_{-0.23}$	$-0.6^{+0.2}_{-0.2}$	$-0.3^{+0.2}_{-0.2}$
Hot Sub-Neptune	$-2.95^{+0.15}_{-0.16}$	$+2.2^{+0.2}_{-0.2}$	$+1.6^{+0.3}_{-0.3}$
Warm Sub-Neptune	$-0.85^{+0.14}_{-0.13}$	$+0.2^{+0.1}_{-0.1}$	$+0.5^{+0.2}_{-0.2}$
Hot Sub-Saturn	$-4.93^{+0.65}_{-0.76}$	$+2.2^{+0.8}_{-0.7}$	$+5.5^{+1.6}_{-1.5}$
Warm Sub-Saturn	$-2.59^{+0.53}_{-0.53}$	$+0.5^{+0.3}_{-0.4}$	$+2.1^{+0.7}_{-0.7}$
Hot Jupiter	$-3.32^{+0.29}_{-0.31}$	$+0.9^{+0.4}_{-0.4}$	$+3.4^{+0.9}_{-0.8}$
Warm Jupiter	…	…	…

Note. Best-fit parameters associated the joint period-metallicity occurrence model defined in Equation (19).

(9%) of our warm super-Earths have [Fe/H] > 0.18 dex, this is consistent with the fact that only 12% of field stars have [Fe/H] > 0.18 dex. As shown in Figures 4 and 10, we do not observe such an absence of warm super-Earths, but instead find nearly equal occurrence of warm super-Earths around stars of wide-ranging metallicities. This underscores the importance of having accurate knowledge of the distribution of field star metallicities in planet-metallicity studies. Nevertheless, accelerated formation of cores in metal-rich disks may explain the rise in warm sub-Neptune occurrence with metallicity.

An alternative to the timing argument, discussed above, is that high disk metallicities may somehow increase the rate of envelope accretion. However, the gas accretion models of Pollack et al. (1996) and (more recently) Lee & Chiang (2015) predict the opposite effect: dusty envelopes should accrete more slowly, due to higher opacities and longer cooling timescales.

As we consider larger warm planets, the metallicity correlation becomes stronger. Sub-Saturns have a β index of $+2.1^{+0.7}_{-0.7}$. If disk metallicity assists with the accretion of gas through accelerated core formation, it may also explain the sub-Saturns. However, abrupt strengthening of the metallicity correlation above 4 $R_\oplus$ and the roughly order of magnitude decrease in the frequency of sub-Saturns compared to sub-Neptunes suggests the formation pathways of sub-Saturns and sub-Neptunes may be quite different.

Around 20 sub-Saturns have well-measured masses from either RVs or TTVs (see Petigura et al. 2017b for a recent compilation). For sub-Saturns, we observe an order of magnitude scatter in the observed masses at a given size, indicating a diversity in core and envelope masses. While some sub-Saturns have ≈ 5 $M_\oplus$ cores, similar to the super-Earths and sub-Neptunes, many have cores of ≈ 50 $M_\oplus$. In addition, the most massive sub-Saturns tend to be found around the most metal-rich hosts (Petigura et al. 2017b).

The existence of ≈ 50 $M_\oplus$ cores in planets with 20% envelope fractions challenges the classic core-accretion models of Pollack et al. (1996) that predict that cores larger than 10 $M_\oplus$ should undergo runaway accretion. Perhaps these massive sub-Saturn cores are the result of the late-stage mergers (or series of mergers) of 10 $M_\oplus$ cores. This formation scenario requires one or more closely spaced 10 $M_\oplus$ planets. As we have shown, the probability for a star to produce a 10 $M_\oplus$ core increases with metallicity. Therefore, the probability for a star to produce two 10 $M_\oplus$ cores likely increases with a steeper power law index. Thus, the production of sub-Saturns by collisions may explain the mass–metallicity dependence and the steeper relationship between metallicity and planet occurrence. This could also explain the mass–metallicity correlation. This theory also predicts that if late-stage mergers play a large role in the formation of sub-Saturns, they should produce relic eccentricities that are observable at later times.

Our planet sample includes only four warm Jupiters, which is insufficient to search for trends. Fischer & Valenti (2005) analyzed 1040 FGK-type stars from the Keck, Lick, and Anglo-Australian Telescope planet search programs and found a planet-metallicity correlation with a $\beta = 2$ index. However,

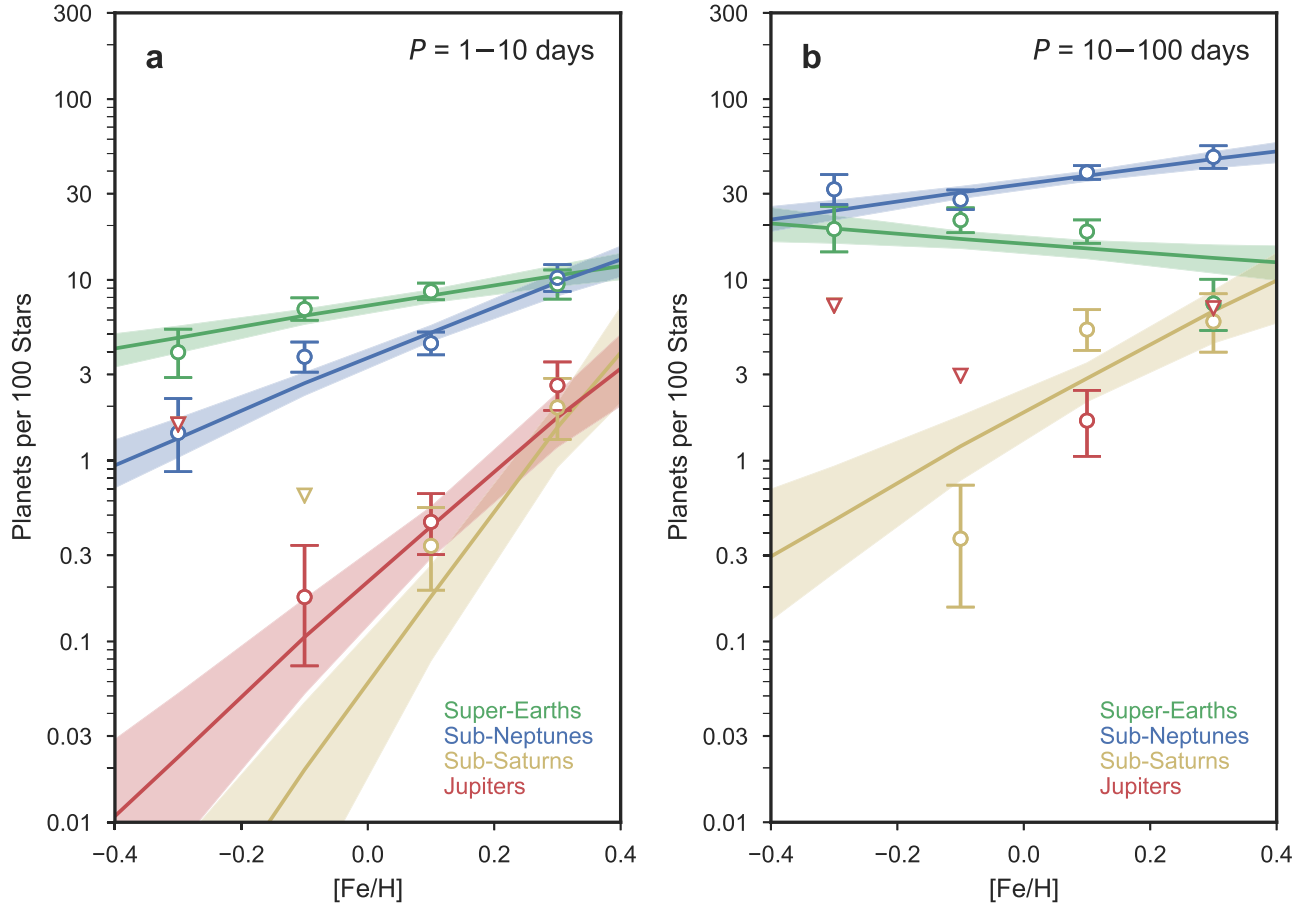

Figure 10. Panel (a): Points show the number of planets per 100 stars for bins of host star metallicity spanning $\Delta M = 0.2$ dex having $P = 1$–10 days. Triangles represent upper limits (90%). The colors correspond to different planet size classes. We modeled the observed occurrence rates with an exponential model, $df \propto P^{\alpha} 10^{\beta M} dM$, where β characterizes the strength of a metallicity correlation (see Section 6, Equation (19)). The solid lines and bands show the best-fitting model and the 1σ credible range of models, respectively. Metallicity correlates with the occurrence of super-Earths ($\beta = +0.6^{+0.2}_{-0.2}$), sub-Neptunes ($\beta = +1.6^{+0.3}_{-0.3}$), sub-Saturns ($\beta = +5.5^{+1.6}_{-1.5}$), and Jupiters ($\beta = +3.4^{+0.9}_{-0.8}$). The strength of the correlation increases with planet size. Panel (b): same as (a) except for planets with $P = 10$–100 days. Warm super-Earths are not correlated with metallicity ($\beta = -0.3^{+0.2}_{-0.2}$). We observe positive correlations for larger planets. For warm sub-Neptunes ($\beta = +0.5^{+0.2}_{-0.2}$); for warm sub-Saturns ($\beta = +2.1^{+0.7}_{-0.7}$). Comparing the two panels, the metallicity correlation is stronger for $P < 10$ days.

this sample included planets with $P = 1$–4000 days, with the bulk of the sample having $P > 300$ days longer than the periods considered here.

7.2. Metallicity and the Prevalence of Short-period Planets

We found that metallicity is associated with the presence of short-period planets having $P < 10$ days. This is consistent with trends observed by Mulders et al. (2016) and Dong et al. (2018), who studied the host star metallicities for samples of 665 and 295 planets, respectively. Both studies used spectroscopically constrained metallicities from LAMOST. Our analysis incorporates the high-precision CKS metallicities for a larger sample of 970 planets.

While close-in planets, of all sizes, are rare compared to more distant planets, their frequency increases as a function of stellar metallicity. In Section 5, when considering just the period distribution of super-Earths and sub-Neptunes, we found that for $P < 10$ days, the occurrence of super-Earths and sub-Neptunes declines with decreasing period. The slope of this falloff is approximated by $df \propto P^{\alpha} d \log P$, with α values of $2.4^{+0.4}_{-0.3}$ and $2.3^{+0.2}_{-0.2}$, respectively.

In Section 6, when considering the joint P–[Fe/H] occurrence distribution, we found that both hot super-Earths and sub-Neptunes are correlated with host star metallicity. This is especially noteworthy for the hot super-Earths, given that the warm super-Earths exhibit no such correlation. For $P < 10$ days, the frequency of hot super-Earths grows from 4% for stars having [Fe/H] = [−0.4, −0.2] dex to 10% for stars having [Fe/H] = [+0.2, +0.4] dex (see Figure 10). We see an even steeper increase in the hot sub-Neptune rates from 1% to 8% for the same metallicity intervals. In Sections 7.2.1 and 7.2.2, we consider two possible formation pathways: (1) in situ models, where planets form near their final location, and (2) high eccentricity migration, where the semimajor axes of fully formed planets are altered by major scattering events. In Section 7.2.3, we consider some observational tests that may distinguish between these scenarios.

7.2.1. In Situ Formation

The prevalence of compact multi-planet systems of super-Earths/sub-Neptunes discovered by *Kepler* helped inspire in situ planet formation models (see, e.g., Hansen & Murray 2013 and Chiang & Laughlin 2013). Chiang & Laughlin (2013) introduced the minimum mass extrasolar nebula (MMESN) as a means of estimating the surface density profile $\Sigma(a)$ of the protoplanetary disk. They took a sample of

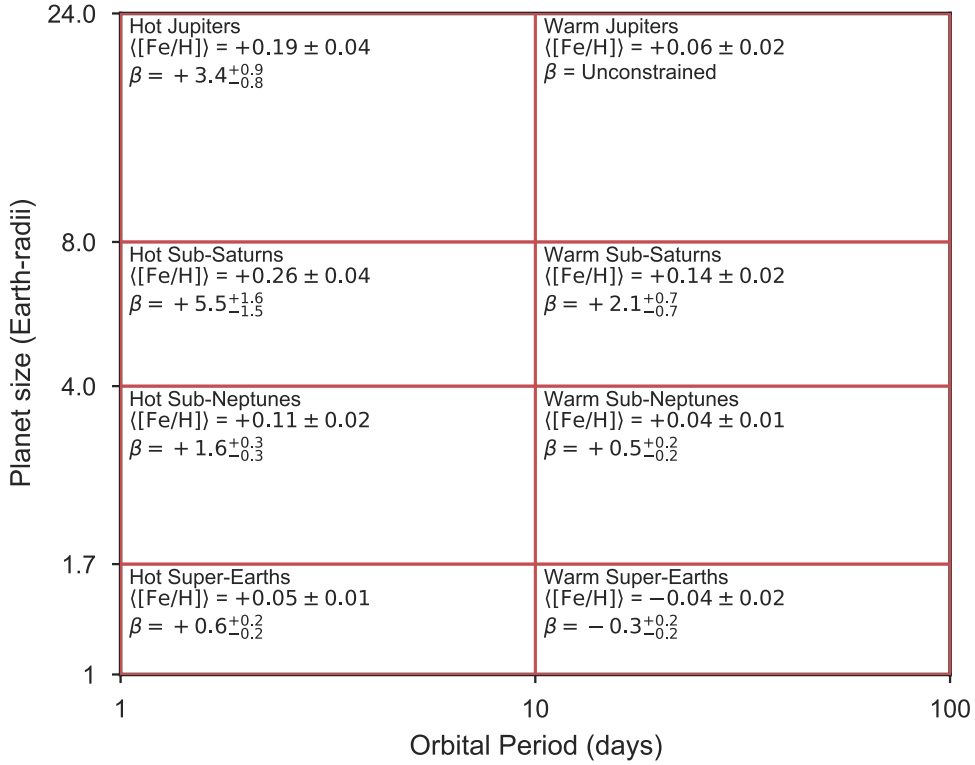

Figure 11. Summary of the planet-metallicity correlation across the P–R_P plane. For each planet subclass studied, we display the average host star metallicity ⟨[Fe/H]⟩ and the β index, which quantifies the strength of the planet-metallicity correlation.

Kepler planets, estimated the mass in solids, and smeared those solids out over an area $2\pi a^2$. The MMESN assumes no large-scale migration of solid building blocks or planets. In the context of the MMESN framework, the falloff in super-Earth and sub-Neptune occurrence at $P < 10$ days reflects a declining surface density profile. Lee & Chiang (2017) offered a physical explanation for declining density profiles. In their models, disks are truncated by interactions with the protostar magnetosphere at the co-rotation radius.

The existence of hot sub-Neptunes presents a challenge to strictly in situ models, because sub-Neptunes with $P < 10$ days are vulnerable to photo-evaporation. Owen & Wu (2017) performed a population synthesis starting with 3–10 $M_\oplus$ cores and a period distribution of $df \propto P^{1.9} d \log P$, which is similar to the power law index presented in this work. They then gave the simulated planets a range of envelope fractions from 1% to 30% and tracked the radii of these planets as they were subjected to photo-evaporation. The Owen & Wu (2017) simulations produced very few sub-Neptunes with $P < 10$ days.

In the Lee & Chiang (2017) model, hot super-Earth/sub-Neptune occurrence rates are set by stellar rotation on the pre-main sequence. It is not clear if stellar metallicity is correlated with pre-main-sequence rotation rates, and Lee & Chiang (2017) did not predict a metallicity dependence for hot super-Earths/sub-Neptunes.

Metallicity may still influence the vertical and radial profiles of the inner disk, even if there is no correlation with stellar rotation. The inner regions of high metallicity disks may have higher densities of free electrons, which could result in stronger coupling to stellar magnetic fields and change the accretion rate and density profile of the disk. The formation of hot super-Earths/sub-Neptunes may also involve significant radial transport of solids, either through the inward drift of "pebbles" that are no longer aerodynamically coupled to the gas disk (see, e.g., Lambrechts & Lega 2017, and references therein). The efficiency of this process is also governed by the radial and vertical density profiles.

Future theoretical modeling of protoplanetary disks is needed to investigate whether in situ models can reproduce the metallicity dependence observed here. In situ models tend to produce dynamically cool planetary systems, which can be corroborated by present-day eccentricity and obliquity measurements

7.2.2. High Eccentricity Migration

High eccentricity migration may explain the hot super-Earth/sub-Neptune period and metallicity distributions. Some of the hot super-Earths/sub-Neptunes may be the result of scattering events between multiple planets. Such encounters preserve the total orbital energy of both planets, but the semimajor axes of both planets may be significantly altered. Dawson & Murray-Clay (2013) proposed a similar mechanism to explain the fact that high eccentricity giant planets with $a = 0.1$–1.0 au, almost exclusively orbit stars with super-solar metallicities.

A formation pathway involving scattering may help to explain the existence of hot sub-Neptunes, which are at risk of photo-evaporative stripping (see Section 7.2.1). Scattering events may occur late enough such that a sub-Neptune is not subjected to the majority its star's XUV output ($t \gtrsim 100$ Myr).

For hot sub-Neptunes, the metallicity correlation is stronger ($\beta = +1.6^{+0.3}_{-0.3}$) than for the warm sub-Neptunes ($\beta = +0.5^{+0.2}_{-0.2}$). This steeper relationship may support a scattering interpretation given that the probability of scattering depends

on the probability of a disk producing two closely spaced planets of comparable mass. If high eccentricity migration is active, the present-day systems planetary systems should be dynamically hot, provided that star–planet interactions are not strong enough to circularize or realign orbits. Again, this may be corroborated with eccentricity and obliquity measurements.

7.2.3. Observational Tests

Additional clues to the formation of hot super-Earths/sub-Neptunes may lie in their orbital eccentricities and degree of alignment with stellar spin axes (obliquities). If planets initially form in the plane of their disks, but are then scattered by other planets, we predict relic eccentricities and obliquities. If these planets form in relative isolation, and are simply a product of the local disk density profile, their eccentricities and obliquities should be low.

Current state-of-the-art RV facilities have provided mass measurements of several dozen hot super-Earths and sub-Neptunes, but are not precise enough to measure orbital eccentricities well. Obliquity measurements via the Rossiter–McLaughlin technique have been made for only a handful of sub-Jovian-size planets (e.g., HAT-P-11, Winn et al. 2010; Hirano et al. 2011). As the precision and observing efficiency of RV facilities improve, we should be able to place tighter constraints on the orbits of these hot planets.

7.3. Hot Jupiters

Hot Jupiters represent an extreme of planet formation, and many theories have been proposed to explain their formation. A non-exhaustive list of theories includes star–planet Kozai (e.g., Wu & Murray 2003), planet–planet Kozai (e.g., Naoz et al. 2011), smooth disk-driven Type-I migration (e.g., Ida & Lin 2008), high eccentricity migration, and in situ formation (e.g., Batygin et al. 2016). Hot Jupiters are also of interest because they offer a point of comparison between the *Kepler* population of planets and those around nearby stars.

We find a hot Jupiter rate of $0.57^{+0.14}_{-0.12}$ %,[21] which is about 1.5σ larger than 0.4 ± 0.1% reported by Howard et al. (2012). This difference cannot be due to differences the treatment of completeness; for hot Jupiters, $p_{\rm det}$ is nearly unity.

The larger occurrence rate presented here is likely due to the improved radius measurements in the CKS sample. Howard et al. (2012) included 13 hot planets with $R_P = 5.6$–$8.0\ R_\oplus$, slightly too small to be classified as a hot Jupiter by their criteria. With our improved CKS planet radii, we have found that these large, hot sub-Saturns are exceptionally rare. Our planet sample $\mathcal{P}$ includes just 3 such planets, while Howard et al. (2012) included 13, even though their stellar parent is less than twice as large as ours $n_{\mathcal{S}}$ (58,041 versus 33,020). Some of the hot sub-Saturns in Howard et al. (2012) were misidentified hot Jupiters.

A major challenge in comparing the hot Jupiter rates in the *Kepler* and the solar neighborhood is their low intrinsic occurrence. Even though our planet sample $\mathcal{P}$ was drawn from a parent stellar sample $\mathcal{S}$ containing 33,020 stars, $\mathcal{P}$ contained just 14 detections due to low intrinsic occurrence and the requirement of transiting geometries. Even blind RV surveys which are not limited to transiting geometries must observe hundreds of stars to detect a few hot Jupiters. Another challenge is that RV surveys are typically not magnitude-limited samples like $\mathcal{S}$. RV surveys often prioritize stars that are single, slowly rotating, and have stellar properties that fall within a preselected range of interest (e.g., M-stars or evolved stars). Wright et al. (2012) attempted to retroactively remove these selection effect from the California Planet Search database, to produce a magnitude-limited sample of stars that could be directly compared to the *Kepler* stellar sample. Wright et al. (2012) reported 10 hot Jupiters from a sample of 836 stars or a rate of 1.2 ± 0.4%, twice the rate presented here. However, the large fractional uncertainty in the Wright et al. (2012) measurement means that these two results differ by only 1.5σ.

In Section 6, we modeled the hot Jupiter occurrence as $df \propto P^{\alpha} 10^{\beta M} d\log P dM$ and found a β index of $+3.4^{+0.9}_{-0.8}$. This strong association between hot Jupiters and metallicity has also been observed by RV stars. Fischer & Valenti (2005) studied the occurrence of Jupiter-mass planets in the CPS sample and modeled their occurrence according to $df \propto 10^{\beta M}$ and β index of 2.0. Johnson et al. (2010) performed a similar analysis, considered an additional mass dependence $df \propto M_\star^{\alpha} 10^{\beta M}$ and found a β index of 1.2 ± 0.2. It is worth recalling that both studies included long period planets ($P > 1$ year), which constituted the bulk of the sample. Guo et al. (2017) repeated the Johnson et al. (2010) analysis, but restricted the study to hot Jupiters and found $\beta = 2.1 \pm 0.7$. While this result differs with our measurement of β by about 2σ, it is clear that hot Jupiters in the solar neighborhood and in the *Kepler* field are both strongly associated with metallicity.

In the previous sections, we have noted the general tendency for the metallically correlation to steepen with decreasing orbital period and increasing planet size, and have hypothesized some processes that could account for these trends such as accelerated core/envelope growth, higher disk densities, and planet–planet scattering. Given this general trend, it is perhaps not surprising that hot Jupiters would have the strongest metallicity correlation. However, given that the hot Jupiters occupy a distinct island on the P–R_P plane, it is possible that their formation pathway is radically different than that of the other planet classes considered here.

7.4. Future Survies

Our target sample contains very few stars (16) with metallicities below -0.4 dex. As a result, we are insensitive to planet occurrence at low metallicities. Probing planet occurrence at low metallicities would shed light on some of the questions raised by this analysis, such as, What is the minimum metallicity needed to form a warm super-Earth?

Since metal-poor stars are rare, a blind survey like *Kepler* must target a large number of stars in order to capture sufficient numbers of low metallicity stars. There is some room for improvement with the existing *Kepler* sample. Our sample was drawn from ≈38,000 of the ≈150,000 stars observed by *Kepler* due to our magnitude limit. Characterizing the metallicities of *Kepler* planet hosts and field stars fainter than $Kp = 14.2$ mag would enlarge the sample of metal-poor stars. However, given that these new stars would be between $Kp = 14$–16 mag, their amenability to the detection of small planets is poor.

Upcoming missions like *TESS* (Ricker et al. 2015) and *PLATO* (Rauer et al. 2014) will survey more stars than *Kepler* and thus cast a significantly larger net for rare metal-poor stars.

[21] The rate of $0.57^{+0.14}_{-0.12}$% is slightly higher than a simple sum of the most likely rates from each hot Jupiter bin from Figure 6 due to the asymmetric uncertainties from low counts per bin.

Gaia will also identify metal-poor stars across the sky. A targeted RV survey of low metallicity stars would also probe planet occurrence at low metallicities. Such a survey would not be restricted to transiting planets, and would need to survey fewer stars to gather statistically significant planet samples.

Finally, one could search for planets in globular clusters. Gilliland et al. (2000) conducted an important early transit search with *HST*, which targeted 47 Tucanae ([Fe/H] = −0.78 dex). The expected yield was 17 hot Jupiters, but the survey yielded no detections, which was attributed to low metallicities. Recently, Masuda & Winn (2017) re-assessed the significance of the null result, using updated knowledge of the typical size, periods, and host star properties of hot Jupiters. Their revised yield calculation was 2 ± 1, and thus the null result does not require a metallicity effect. However, based on the steep dependence of hot Jupiter occurrence with metallicity, yields would likely be low. Future globular cluster surveys with higher precision or longer baselines would help constrain planet population at −1 dex. However, high stellar densities of globulars may complicate direct comparisons to field stars.

8. Conclusion

We have measured planet occurrence as a function of period, radius, and stellar metallicity within a magnitude-limited sample of *Kepler* target stars. We have leveraged precise planet radii and stellar metallicities from the CKS catalog of *Kepler* planet-hosting stars, as well as the metallicity distribution of *Kepler* field stars measured by LAMOST. In summary, for $P < 100$ days the default planetary system contains either no planets detectable by *Kepler* (e.g., the solar system) or a system of one or more super-Earths/sub-Neptunes with $P = 10$–100 days. Stars with high metallicity are associated with some mechanism(s) that also allows for "misplaced" planets: sub-Saturns/Jupiters with $P < 100$ days and super-Earths/sub-Neptunes with $P < 10$ days. We have noted the general features of planet occurrence as a function of P, R_P, and [Fe/H], and have offered some speculation regarding their physical origins. Detailed population synthesis models that treat the disk profile, growth of cores, migration, dynamical instabilities, and photo-evaporation must produce these features.

The CKS project was conceived, planned, and initiated by A.W.H., G.W.M., J.A.J., H.T.I., and T.D.M. A.W.H., G.W.M., J.A.J. acquired Keck telescope time to conduct the magnitude-limited survey. We thank the many observers who contributed to the measurements reported here.

We thank the referee, Lars Buchhave, for his detailed and thoughtful comments on the techniques and interpretations. We also thank Konstantin Batygin, Brendan Bowler, Ian Crossfield, and Eve Lee for enlightening conversations that improved the final manuscript.

Kepler was competitively selected as the tenth NASA Discovery mission. Funding for this mission is provided by the NASA Science Mission Directorate. We thank the Kepler Science Office, the Science Operations Center, the Threshold Crossing Event Review Team (TCERT), and the Follow-up Observations Program (FOP) Working Group for their work on all steps in the planet discovery process, ranging from selecting target stars and pointing the Kepler telescope to developing and running the photometric pipeline to curating and refining the catalogs of Kepler planets.

The Guoshoujing Telescope (the Large Sky Area Multi-Object Fiber Spectroscopic Telescope, LAMOST) is a National Major Scientific Project built by the Chinese Academy of Sciences. Funding for the project has been provided by the National Development and Reform Commission. LAMOST is operated and managed by the National Astronomical Observatories, Chinese Academy of Sciences.

E.A.P. acknowledges support from Hubble Fellowship grant HST-HF2-51365.001-A, awarded by the Space Telescope Science Institute, which is operated by the Association of Universities for Research in Astronomy, Inc., for NASA under contract NAS 5-26555. L.M.W. acknowledges support from Gloria and Ken Levy and from the Trottier Family. T.D.M. acknowledges NASA grant NNX14AE11G.

This work made use of NASA's Astrophysics Data System Bibliographic Services.

Finally, the authors wish to recognize and acknowledge the very significant cultural role and reverence that the summit of Maunakea has always had within the indigenous Hawaiian community. We are most fortunate to have the opportunity to conduct observations from this mountain.

Software: All code used in this paper is available at https://github.com/California-Planet-Search/cksmet/. We made use of the following publicly available Python modules: astropy (Astropy Collaboration et al. 2013), isoclassify (Huber et al. 2017), lmfit (Newville et al. 2014), matplotlib (Hunter 2007), numpy/scipy (van der Walt et al. 2011), pandas (McKinney 2010), and pinky (https://github.com/pgromano/pinky).

Appendix A
Calibration of LAMOST Metallicities

This paper relies on CKS metallicities for the planet-hosting stars and LAMOST metallicities to characterize the metallicity distribution of *Kepler* field stars. Differences in metallicity scales often result from different spectral data sets or spectroscopic pipelines. The CKS and LAMOST surveys used different spectra, line-lists, and model-atmospheres. Here, we assess the agreement between the CKS and LAMOST metallicity scales, correct for a small offset, and report the uncertainty in our calibration.

There are 476 stars observed by both CKS and LAMOST. Figure 12 shows the difference between LAMOST and CKS metallicities (Δ[Fe/H]) as a function of CKS metallicity, *Kp*, and the S/N of the LAMOST spectra. On average, the LAMOST metallicities are 0.04 dex lower, and we observe a small metallicity-dependent systematic trend. The dispersion of Δ[Fe/H] increases for fainter stars or decreasing LAMOST S/N (the CKS spectra have homogeneous S/N). This indicates that the precision of the LAMOST metallicities declines with decreasing S/N.

We derived a correction Δ that calibrates the LAMOST metallicities onto the CKS scale via $\mathrm{LAMOST_{cal}} = \mathrm{LAMOST_{raw}} + \Delta$. The correction is a linear function of the following form:

$$\Delta[\mathrm{Fe/H}] = c_0 + c_1\left(\frac{[\mathrm{Fe/H}]}{0.1\ \mathrm{dex}}\right). \quad (22)$$

The coefficients are chosen such that they minimize the rms difference between the calibrated LAMOST and CKS parameters. Before fitting, we removed 17 stars where the CKS

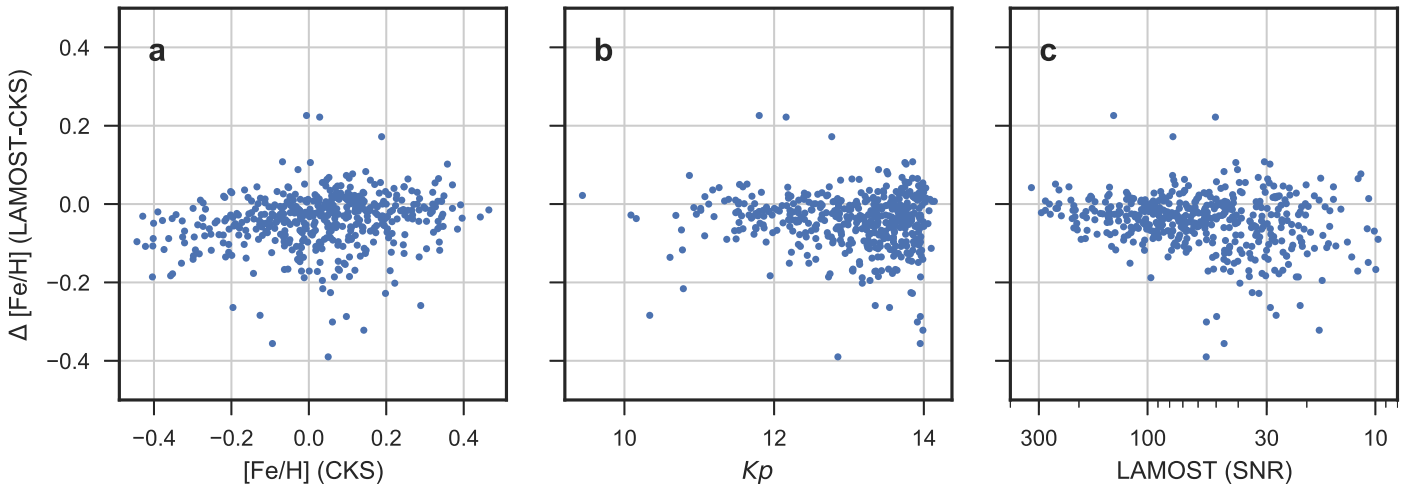

Figure 12. Difference between CKS and LAMOST [Fe/H] for 476 stars in common as a function of CKS metallicity (a), *Kp* (b), and the reported S/N of the LAMOST spectra (c). We observe a small metallicity-dependent systematic difference, which we calibrate out in Appendix A.

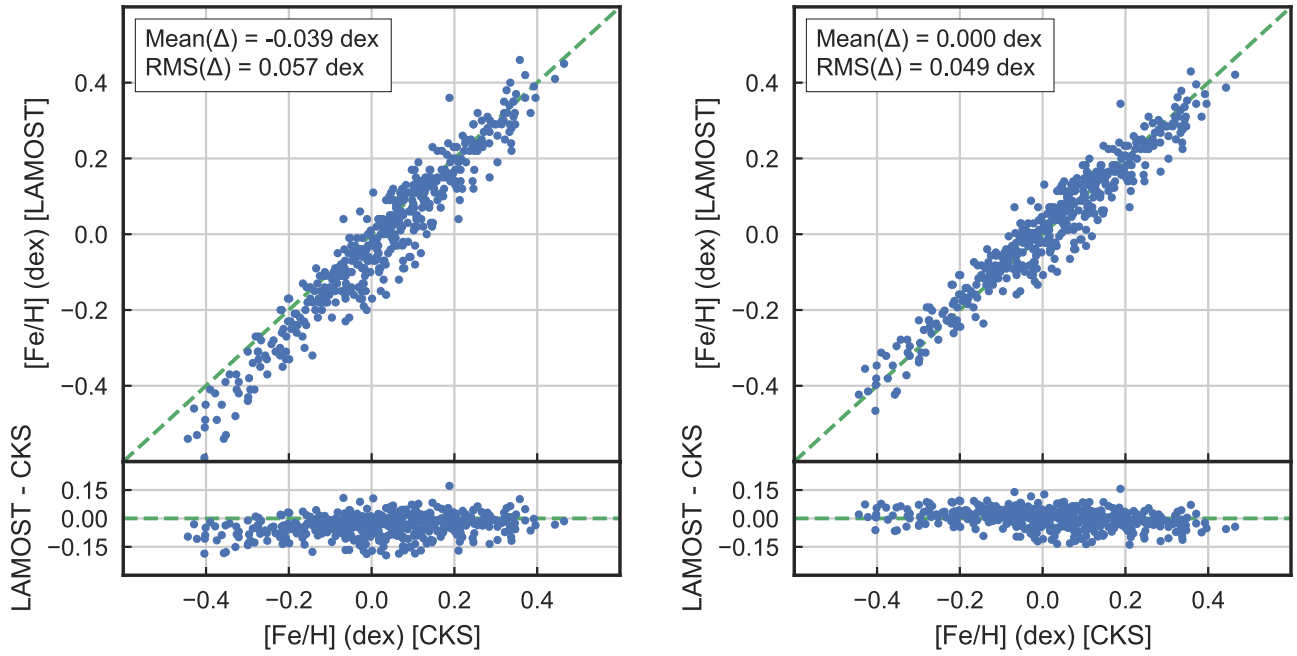

Figure 13. Comparison of CKS and LAMOST metallicities. (Left) Uncalibrated LAMOST metallicities as a function of CKS metallicities for stars in common. The green dashed line represents equality. On average, the LAMOST metallicities are 0.05 dex lower than the CKS values. (Right) comparison of CKS and LAMOST metallicities after removing a linear systematic trend.

and LAMOST metallicities differed by more than 0.2 dex. These outliers are likely due to rare failure modes of the LAMOST pipeline.[22]

We estimated the uncertainties on the best-fit parameters using bootstrap resampling (with replacement), as described in Press (2002). The best-fit coefficients are $c_0 = 0.037 \pm 0.002$ and $c_1 = -0.015 \pm 0.001$. We show a one-to-one comparison of the CKS and LAMOST metallicities (without the 17 outliers) in Figure 13. After placing the LAMOST metallicities on the CKS scale, there is no residual mean offset (by construction) and a rms dispersion of 0.05 dex.

[22] To verify that our calibration does not depend sensitively on our choice to remove outliers, we left them in but solved for the coefficients that minimized the sum of the absolute differences, a metric which is resistant to outliers. The coefficients agreed to 1.5σ or better.

Our calibration amounts to a correction of $+0.110 \pm 0.007$ dex at [Fe/H] $= -0.5$ dex and -0.036 ± 0.007 dex at [Fe/H] $= +0.5$ dex. The uncertainty associated with our correction at -0.5 and $+0.5$ dex sets an upper bound on the residual systematic offset between the LAMOST and CKS metallicity scales relevant to this paper, which treats planet hosts having metallicities between -0.5 and $+0.5$ dex. We estimate that the CKS and the calibrated LAMOST metallicities are consistent to 0.01 dex.

To verify that the main results of this paper are not sensitive to zero-point offsets between the CKS and LAMOST metallicities scales, we performed a parallel analysis after perturbing our calibrated LAMOST metallicities by 0.01 dex. We compared the slopes of the metallicity distributions for different planet classes, which is quantified by the β index in Equation (19). Adding 0.01 dex to the LAMOST metallicities

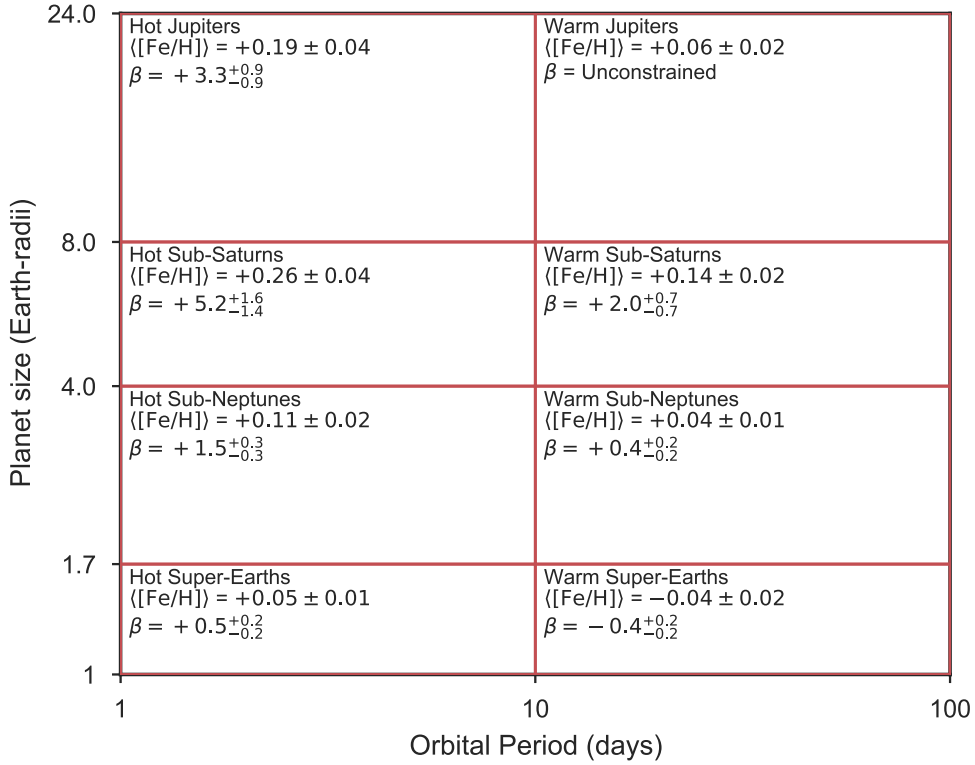

Figure 14. Same as Figure 11, but after adding 0.01 dex to the LAMOST metallicities to simulate the effect of residual offsets between the CKS and LAMOST metallicity scales. For all planet classes, the change in β is smaller than the adopted uncertainty on β, which incorporates counting statistics alone. The uncertainties on β are therefore dominated by counting statistics, rather than uncertainties in our CKS-LAMOST metallicity calibration.

increases $n_{\rm trial}$ in the high metallicity bins, which results in lower measured occurrence at high metallicities. For low metallicity bins, $n_{\rm trial}$ decreases, resulting larger measured occurrence. The net effect is that the slope of the metallicity distribution becomes less steep (i.e., smaller β indices). (Subtracting 0.01 dex from the calibrated LAMOST metallicities has the opposite effect and results in larger beta-indices.)

Figure 14 is analogous to Figure 11, but we created it after shifting the LAMOST metallicities by +0.01 dex. Comparing the β indices, these two figures summarize the effects of a possible offset in metallicity scales. For all planet classes, β decreases, as expected. For example, for warm super-Earths, β decreases from -0.3 ± 0.2 to -0.4 ± 0.2; for warm sub-Neptunes, β decreases from $+0.5 \pm 0.2$ to $+0.4 \pm 0.2$. For all planet classes, the change in β is smaller than our adopted 1σ uncertainties from our maximum-likelihood fitting, which only incorporates counting statistics. Therefore, we conclude that the uncertainties on the β indices due to our LAMOST-CKS calibration are smaller than those due to counting statistics. Thus, the conclusions of this paper are not sensitive to errors in our LAMOST-CKS metallicity calibration.

Appendix B
Metallicity and Planet Detectability

In Section 3, we compared the properties of planets belonging to host stars of different metallicities, and in Section 6, we computed planet occurrence as a function of host star metallicity. For both calculations, we made the assumption that planet detectability is not a strong function of host star metallicity. Here, we check the validity of such an assumption.

Metallicity could correlate with planet detectability through a correlation with stellar size or noise properties. As a direct measure of the possible dependence of planet detectability with metallicity, we computed the single transit signal-to-noise ratio (S/N_1) of a 3-hour transit of putative 1 $R_\oplus$ planet transiting each star in the LAMOST sample. S/N_1 was computed according to

$$\mathrm{SNR}_1 = \left(\frac{R_P}{R_\star}\right)^2 \left(\frac{1}{\mathrm{CDPP3}}\right), \tag{23}$$

where $R_P = 1\ R_\oplus$ and $R_\star$ was computed using LAMOST $T_{\rm eff}$, $\log g$, and [Fe/H] measurements and the publicly available `isoclassify` package (Huber et al. 2017).[23] As Figure 15 shows, we do not observe a significant dependence of planet detectability with stellar metallicity.

Appendix C
Planet Occurrence: Period and Radius

Here we provide two tables to supplement Section 5, which treats planet occurrence as a function of period and radius. Table 8 lists the occurrence measurements displayed in Figure 6 along with upper limits. Table 9 is a sampling of the occurrence distribution shown in Figure 8. We sample the occurrence the where the following conditions are met:

1. P = 1–300 days (i.e., the range of periods shown in Figure 8).

[23] https://github.com/danxhuber/isoclassify

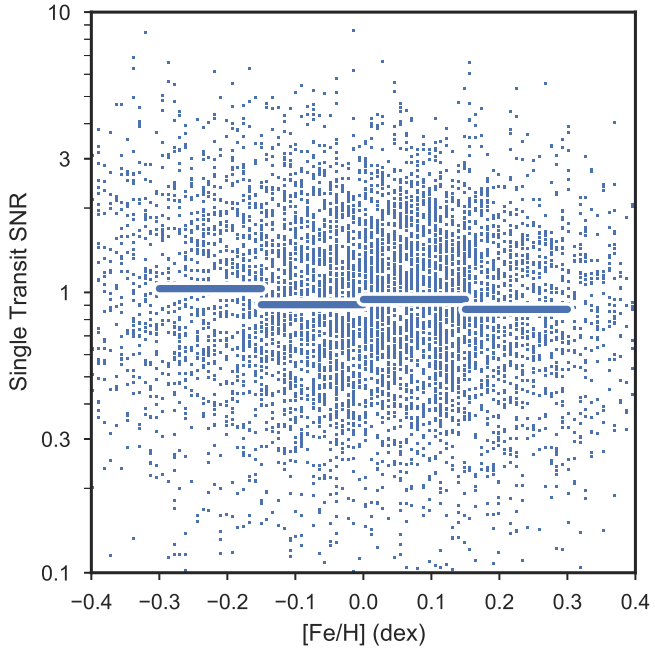

Figure 15. Planet detectability vs. metallicity. The *y*-axis shows the signal-to-noise ratio of a putative 1 $R_\oplus$ planet, which has a transit duration of three hours. We do not observe a significant trend in planet detectability with metallicity.

Table 8
Planet Occurrence

$R_{P,1}$	$R_{P,2}$	P_1	P_2	$n_{\rm pl}$	$p_{\rm det}$	$n_{\rm trial}$	$f_{\rm cell}$
1.00	1.41	1.00	1.78	7	0.95	6551.6	$0.12^{+0.05}_{-0.04}$
1.00	1.41	1.78	3.16	15	0.93	4369.1	$0.36^{+0.10}_{-0.08}$
1.00	1.41	3.16	5.62	40	0.89	2874.8	$1.41^{+0.23}_{-0.21}$
1.00	1.41	5.62	10.00	48	0.84	1847.5	$2.63^{+0.39}_{-0.36}$
1.00	1.41	10.00	17.78	51	0.76	1144.3	$4.50^{+0.64}_{-0.59}$
1.00	1.41	17.78	31.62	23	0.66	673.5	$3.50^{+0.76}_{-0.65}$
1.00	1.41	31.62	56.23	10	0.53	370.5	$2.87^{+0.95}_{-0.78}$
1.00	1.41	56.23	100.00	4	0.39	187.4	$2.49^{+1.26}_{-0.97}$
1.00	1.41	100.00	177.83	0	0.27	86.4	<2.61
1.00	1.41	177.83	316.23	0	0.16	36.3	…

Note. Planet occurrence computed over various intervals in the P–R_P plane spanning 0.25 dex and 0.15 dex, respectively. Each bin boundary is given by $[R_{P,1}, R_{P,2}]$ and $[P_1, P_2]$ respectively. We list the number of planets per bin $n_{\rm pl}$, the pipeline detectability $p_{\rm det}$, and the number of effective trials $n_{\rm trial}$. The number of planets per 100 stars per bin is given by $f_{\rm cell}$, which is also shown graphically in Figure 6. We report 90% upper limits on $f_{\rm cell}$ when there are no planets in a bin, and we do not report $f_{\rm cell}$ when $p_{\rm det} < 0.25$.

(This table is available in its entirety in machine-readable form.)

2. $R_P = 0.5$–32 $R_\oplus$ (i.e., the range of planet sizes shown in Figure 8).
3. $\log(R_P/R_\oplus) > 0.15 \log(P/1\ \text{day})$, which ensures that pipeline completeness, $p_{\rm det} > 0.25$.

The integrated occurrence within this domain is 110.7 planets per 100 stars. We simulated the periods and radii of a population of 110,733 planets in a sample of 100,000 Sun-like stars by drawing 110,733 (P, R_P) pairs according to their measured occurrence rates using the Python package pinky. These samples are listed in Table 9.

Table 9
Simulated Planet Properties

Planet	P days	R_P $R_\oplus$
0	45.194	1.185
1	94.327	0.964
2	56.235	1.100
3	44.367	2.254
4	26.778	0.827
5	67.644	6.516
6	28.391	1.366
7	236.817	7.830
8	336.385	11.736
9	22.677	2.457

Note. Simulated periods and radii of 110,733 planets in a population of 100,000 stars based on the measured occurrence rates from *Kepler* (Section 5). This table may be used to compute yield simulations for future surveys or integrated occurrence values over arbitrary bins of P and R_P.

(This table is available in its entirety in machine-readable form.)

ORCID iDs

Erik A. Petigura https://orcid.org/0000-0003-0967-2893
Geoffrey W. Marcy https://orcid.org/0000-0002-2909-0113
Joshua N. Winn https://orcid.org/0000-0002-4265-047X
Lauren M. Weiss https://orcid.org/0000-0002-3725-3058
Benjamin J. Fulton https://orcid.org/0000-0003-3504-5316
Andrew W. Howard https://orcid.org/0000-0001-8638-0320
Evan Sinukoff https://orcid.org/0000-0002-5658-0601
Howard Isaacson https://orcid.org/0000-0002-0531-1073
Timothy D. Morton https://orcid.org/0000-0002-8537-5711

References

Adams, E. R., Ciardi, D. R., Dupree, A. K., et al. 2012, AJ, 144, 42
Adams, E. R., Dupree, A. K., Kulesa, C., & McCarthy, D. 2013, AJ, 146, 9
Akeson, R. L., Chen, X., Ciardi, D., et al. 2013, PASP, 125, 989
Astropy Collaboration, Robitaille, T. P., Tollerud, E. J., et al. 2013, A&A, 558, A33
Baranec, C., Ziegler, C., Law, N. M., et al. 2016, AJ, 152, 18
Barclay, T., Rowe, J. F., Lissauer, J. J., et al. 2013, Natur, 494, 452
Batygin, K., Bodenheimer, P. H., & Laughlin, G. P. 2016, ApJ, 829, 114
Bowler, B. P., Liu, M. C., Shkolnik, E. L., & Tamura, M. 2015, ApJS, 216, 7
Brown, T. M., Latham, D. W., Everett, M. E., & Esquerdo, G. A. 2011, AJ, 142, 112
Buchhave, L. A., Bizzarro, M., Latham, D. W., et al. 2014, Natur, 509, 593
Buchhave, L. A., & Latham, D. W. 2015, ApJ, 808, 187
Buchhave, L. A., Latham, D. W., Johansen, A., et al. 2012, Natur, 486, 375
Burke, C. J., Christiansen, J. L., Mullally, F., et al. 2015, ApJ, 809, 8
Cartier, K. M. S., Gilliland, R. L., Wright, J. T., & Ciardi, D. R. 2015, ApJ, 804, 97
Catanzarite, J., & Shao, M. 2011, ApJ, 738, 151
Chen, H., & Rogers, L. A. 2016, ApJ, 831, 180
Chiang, E., & Laughlin, G. 2013, MNRAS, 431, 3444
Christiansen, J. L., Clarke, B. D., Burke, C. J., et al. 2015, ApJ, 810, 95
Cui, X.-Q., Zhao, Y.-H., Chu, Y.-Q., et al. 2012, RAA, 12, 1197
Cumming, A., Butler, R. P., Marcy, G. W., et al. 2008, PASP, 120, 531
Dawson, R. I., Chiang, E., & Lee, E. J. 2015, MNRAS, 453, 1471
Dawson, R. I., & Murray-Clay, R. A. 2013, ApJL, 767, L24
De Cat, P., Fu, J. N., Ren, A. B., et al. 2015, ApJS, 220, 19
Dong, S., Xie, J.-W., Zhou, J.-L., Zheng, Z., & Luo, A. 2018, PNAS, 115, 266
Dong, S., Zheng, Z., Zhu, Z., et al. 2014, ApJL, 789, L3
Dong, S., & Zhu, Z. 2013, ApJ, 778, 53
Dressing, C. D., Adams, E. R., Dupree, A. K., Kulesa, C., & McCarthy, D. 2014, AJ, 148, 78
Dressing, C. D., & Charbonneau, D. 2015, ApJ, 807, 45

Everett, M. E., Barclay, T., Ciardi, D. R., et al. 2015, AJ, 149, 55
Fischer, D. A., & Valenti, J. 2005, ApJ, 622, 1102
Foreman-Mackey, D., Hogg, D. W., Lang, D., & Goodman, J. 2013, PASP, 125, 306
Foreman-Mackey, D., Hogg, D. W., & Morton, T. D. 2014, ApJ, 795, 64
Fressin, F., Torres, G., Charbonneau, D., et al. 2013, ApJ, 766, 81
Fulton, B. J., Petigura, E. A., Howard, A. W., et al. 2017, AJ, 154, 109
Furlan, E., Ciardi, D. R., Everett, M. E., et al. 2017, AJ, 153, 71
Ghezzi, L., Cunha, K., Smith, V. V., et al. 2010, ApJ, 720, 1290
Gilliland, R. L., Brown, T. M., Guhathakurta, P., et al. 2000, ApJL, 545, L47
Gilliland, R. L., Cartier, K. M. S., Adams, E. R., et al. 2015, AJ, 149, 24
Gonzalez, G. 1997, MNRAS, 285, 403
Goodman, J., & Weare, J. 2010, Communications in Applied Mathematics and Computational Science, 5, 65
Guo, X., Johnson, J. A., Mann, A. W., et al. 2017, ApJ, 838, 25
Hadden, S., & Lithwick, Y. 2017, AJ, 154, 5
Hansen, B., & Murray, N. 2013, ApJ, 775, 53
Hirano, T., Narita, N., Shporer, A., et al. 2011, PASJ, 63, 531
Horch, E. P., Howell, S. B., Everett, M. E., & Ciardi, D. R. 2012, AJ, 144, 165
Horch, E. P., Howell, S. B., Everett, M. E., & Ciardi, D. R. 2014, ApJ, 795, 60
Howard, A. W., Marcy, G. W., Bryson, S. T., et al. 2012, ApJS, 201, 15
Howell, S. B., Everett, M. E., Sherry, W., Horch, E., & Ciardi, D. R. 2011, AJ, 142, 19
Huber, D., Zinn, J., Bojsen-Hansen, M., et al. 2017, ApJ, 844, 102
Hunter, J. D. 2007, CSE, 9, 90
Ida, S., & Lin, D. N. C. 2008, ApJ, 673, 487
Jenkins, J. M., Caldwell, D. A., Chandrasekaran, H., et al. 2010, ApJL, 713, L87
Jin, S., Mordasini, C., Parmentier, V., et al. 2014, ApJ, 795, 65
Johnson, J. A., Aller, K. M., Howard, A. W., & Crepp, J. R. 2010, PASP, 122, 905
Johnson, J. A., Petigura, E. A., Fulton, B. J., et al. 2017, AJ, 154, 108
Lambrechts, M., & Lega, E. 2017, A&A, 606, A146
Law, N. M., Morton, T., Baranec, C., et al. 2014, ApJ, 791, 35
Lee, E. J., & Chiang, E. 2015, ApJ, 811, 41
Lee, E. J., & Chiang, E. 2017, ApJ, 842, 40
Lillo-Box, J., Barrado, D., & Bouy, H. 2012, A&A, 546, A10
Lillo-Box, J., Barrado, D., & Bouy, H. 2014, A&A, 566, A103
Lissauer, J. J. 1995, Icar, 114, 217
Lopez, E. D., & Fortney, J. J. 2013, ApJ, 776, 2
Luo, A.-L., Zhang, H.-T., Zhao, Y.-H., et al. 2012, RAA, 12, 1243
Luo, A.-L., Zhao, Y.-H., Zhao, G., et al. 2015, RAA, 15, 1095
Marcy, G. W., Isaacson, H., Howard, A. W., et al. 2014, ApJS, 210, 20
Masuda, K., & Winn, J. N. 2017, AJ, 153, 187
Mathur, S., Huber, D., Batalha, N. M., et al. 2017, ApJS, 229, 30
Mazeh, T., Holczer, T., & Faigler, S. 2016, A&A, 589, A75
McKinney, W. 2010, in Proc. 9th Python in Science Conf., ed. S. van der Walt & J. Millman, 51
Morton, T. D. 2015, isochrones: Stellar model grid package, Astrophysics Source Code Library, ascl:1503.010
Mulders, G. D., Pascucci, I., & Apai, D. 2015, ApJ, 798, 112
Mulders, G. D., Pascucci, I., Apai, D., Frasca, A., & Molenda-Żakowicz, J. 2016, AJ, 152, 187
Mullally, F., Coughlin, J. L., Thompson, S. E., et al. 2015, ApJS, 217, 31
Naoz, S., Farr, W. M., Lithwick, Y., Rasio, F. A., & Teyssandier, J. 2011, Natur, 473, 187
Newville, M., Stensitzki, T., Allen, D. B., & Ingargiola, A. 2014, LMFIT: Non-Linear Least-Square Minimization and Curve-Fitting for Python, Zenodo, doi:10.5281/zenodo.11813
Nordström, B., Mayor, M., Andersen, J., et al. 2004, A&A, 418, 989
Owen, J. E., & Wu, Y. 2013, ApJ, 775, 105
Owen, J. E., & Wu, Y. 2017, ApJ, 847, 29
Petigura, E. A. 2015, PhD thesis, Univ. California
Petigura, E. A., Howard, A. W., & Marcy, G. W. 2013, PNAS, 110, 19273
Petigura, E. A., Howard, A. W., Marcy, G. W., et al. 2017a, AJ, 154, 107
Petigura, E. A., Sinukoff, E., Lopez, E. D., et al. 2017b, AJ, 153, 142
Pollack, J. B., Hubickyj, O., Bodenheimer, P., et al. 1996, Icar, 124, 62
Press, W. H. (ed.) 2002, Numerical Recipes in C++: the Art of Scientific Computing (Cambridge: Cambridge Univ. Press)
Rauer, H., Catala, C., Aerts, C., et al. 2014, ExA, 38, 249
Ricker, G. R., Winn, J. N., Vanderspek, R., et al. 2015, JATIS, 1, 014003
Rogers, L. A. 2015, ApJ, 801, 41
Santos, N. C., Israelian, G., & Mayor, M. 2004, A&A, 415, 1153
Schlaufman, K. C. 2015, ApJL, 799, L26
Sinukoff, E., Howard, A. W., Petigura, E. A., et al. 2017, AJ, 153, 271
Smalley, B., Anderson, D. R., Collier-Cameron, A., et al. 2012, A&A, 547, A61
Sousa, S. G., Santos, N. C., Mayor, M., et al. 2008, A&A, 487, 373
Sullivan, P. W., Winn, J. N., Berta-Thompson, Z. K., et al. 2015, ApJ, 809, 77
Traub, W. A. 2012, ApJ, 745, 20
van der Walt, S., Colbert, S. C., & Varoquaux, G. 2011, CSE, 13, 22
Vogt, S. S., Allen, S. L., Bigelow, B. C., et al. 1994, Proc. SPIE, 2198, 362
Wang, J., & Fischer, D. A. 2015, AJ, 149, 14
Wang, J., Fischer, D. A., Horch, E. P., & Xie, J.-W. 2015a, ApJ, 806, 248
Wang, J., Fischer, D. A., Xie, J.-W., & Ciardi, D. R. 2015b, ApJ, 813, 130
Weiss, L. M., & Marcy, G. W. 2014, ApJL, 783, L6
Winn, J. N., Johnson, J. A., Howard, A. W., et al. 2010, ApJL, 723, L223
Wright, J. T., Marcy, G. W., Howard, A. W., et al. 2012, ApJ, 753, 160
Wu, Y., & Murray, N. 2003, ApJ, 589, 605
Youdin, A. N. 2011, ApJ, 742, 38
Zapolsky, H. S., & Salpeter, E. E. 1969, ApJ, 158, 809
Zhao, G., Zhao, Y.-H., Chu, Y.-Q., Jing, Y.-P., & Deng, L.-C. 2012, RAA, 12, 723